



**US Army Corps
of Engineers®**

Engineer Research and
Development Center

Methods for Field Studies of the Effects of Military Smokes, Obscurants, and Riot-control Agents on Threatened and Endangered Species

Volume 3: Statistical Methods

Debra M. Cassels, Anthony J. Krzysik,
and Keturah A. Reinbold

September 2001

20011016 070

Methods for Field Studies of the Effects of Military Smokes, Obscurants, and Riot-control Agents on Threatened and Endangered Species

Volume 3: Statistical Methods

by Debra M. Cassels, Anthony J. Krzysik, and Keturah A. Reinbold

U.S. Army Engineer Research and Development Center
Construction Engineering Research Laboratory
PO Box 9005
Champaign, IL 61826-9005

Final report

Approved for public release; distribution is unlimited.

Volume 1 of this report series will be an overview of methods examined in the series, their application, and applicable regulations. Volume 2 (CERL Technical Report 97/140, September 1997) reviews methods for assessing ecological risks. Volume 3 (this report) discusses strategies for developing a statistically sound approach to assessing the effects of military smokes, obscurants, and riot-control agents. Volume 4 (CERL Technical Report 99/56, July 1999) examines chemical analytical methods for isolating and detecting the components of smokes, obscurants, and riot-control agents from environmental media.

Foreword

This study was conducted for the Strategic Environmental Research and Development Program (SERDP) under project number CS-766, "Identification, Assessment, and Mitigation of Impacts of Military Related Chemicals and Pollutants on TES" and project number CS-507, "Threatened, Endangered, and Sensitive Resources: Impact of Smokes and Obscurants on TES." Congress established SERDP through Public Law 101-510 on 5 November 1990 (10 U.S.C. 2901-2904). SERDP is a joint multi-agency (Department of Defense [DoD], Department of Energy [DOE], and Environmental Protection Agency [EPA]) effort

to support environmental quality research, development, demonstration, and applications programs. The technical monitor at the beginning of this work was Dr. Femi A. Ayorinde, Cleanup and Conservation Program Manager, SERDP.



Additional funding was provided by the Office of the Directorate of Environmental Programs (DAIM), Assistant Chief of Staff (Installation Management) (ACSIM) under 622720896, Base Facilities Environmental Quality, Work Unit A896-LL-TC6, "Threshold Impacts for Smokes and Obscurants on TES." The technical monitor was Phil Pierce (DAIM-ED-N).

The work was performed by the Ecological Processes Branch (CN-N) of the Installations Division (CN), Construction Engineering Research Laboratory (CERL). The CERL Principal Investigator was Dr. Keturah A. Reinbold. Debra M. Cassels was an employee of the University of Illinois, Urbana-Champaign, assigned to CERL under an Intergovernmental Personnel Mobility Agreement,

DISCLAIMER

The contents of this report are not to be used for advertising, publication, or promotional purposes. Citation of trade names does not constitute an official endorsement or approval of the use of such commercial products. All product names and trademarks cited are the property of their respective owners.

The findings of this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

DESTROY THIS REPORT WHEN IT IS NO LONGER NEEDED. DO NOT RETURN IT TO THE ORIGINATOR.

and Anthony J. Krzysik was a CERL researcher. Patricia M. Kirby, Colorado State University contractor, coordinated preparation of the final report. The technical editor was Linda L. Wheatley, Information Technology Laboratory — Champaign. Dr. Harold E. Balbach, CN-N, is the Project Leader for threatened and endangered species research. Stephen E. Hodapp is Chief, CN-N, and Dr. John T. Bandy is Chief, CN. The associated Technical Director is Dr. William D. Severinghaus. The Acting Director of CERL is Dr. Alan W. Moore.

CERL is an element of the U.S. Army Engineer Research and Development Center (ERDC), U.S. Army Corps of Engineers. The Director of ERDC is Dr. James R. Houston and the Deputy Commander is A.J. Roberto, Jr.

Contents

Foreword.....	1
1 Introduction.....	7
Background	7
Objectives.....	7
Approach	8
Scope	8
Mode of Technology Transfer	9
2 Sampling Design Considerations	10
Introduction.....	10
Purpose: Classification of Field Research — Mensurative and Manipulative Studies	12
<i>Introduction.....</i>	12
<i>Desired Outcome of the Investigation.....</i>	13
<i>Population of Concern.....</i>	13
<i>Parameters of Interest.....</i>	13
<i>Facts Already Known About the Situation or Problem</i>	14
<i>Assumptions Needed To Initiate the Investigation</i>	15
<i>Basic Nature of the Problem: Research, Inventory, Monitoring, or Conformance.....</i>	16
<i>Temporal Nature of the Problem: One-time, Short-range, or Long-range.....</i>	16
<i>Spatial Nature of the Problem: Local, Regional, or Global.....</i>	16
Question: Definition of Objectives	17
Hypothesis: Selection of Correct Conceptual, Estimation, and Predictive Models for Smoke Effects on T&E Species	19
<i>Definition.....</i>	19
<i>Conceptual Models.....</i>	21
<i>Estimation Models</i>	22
<i>Predictive Models</i>	23
Sampling Design: Development of Appropriate Strategies for Allocating Treatments and Collecting Samples	23
<i>Determination of True Population To Be Sampled</i>	24
<i>Special Considerations for Sampling Abiotic and Biotic Media.....</i>	25
<i>Identification of Confounding Factors</i>	31
<i>Selection of Appropriate Variables.....</i>	32
<i>Selection of Appropriate Experimental Units.....</i>	34

<i>Systematic Versus Random Sampling</i>	36
<i>Nonhomogeneous Mixing and Deposition of Chemicals in the Environment</i>	38
<i>Independence of Experimental Units</i>	38
<i>Sampling Units</i>	38
<i>Types of Designs</i>	39
Execution of Experiment: Sample Collection and Analysis	41
<i>Pilot Study</i>	41
<i>Quality Control</i>	42
<i>Number of Samples</i>	43
<i>Significance Level (α) and Statistical Power ($1 - \beta$)</i>	43
<i>P-Values</i>	45
<i>Physical Size (Volume or Mass) of Sample Unit</i>	46
<i>Length of Sampling Period</i>	46
3 Statistical Analysis Considerations	47
Data Types and Data Quality.....	47
<i>Data Types</i>	47
<i>Data Quality</i>	49
Approaches to Statistical Analysis.....	51
<i>Estimation</i>	52
<i>Descriptive Statistics</i>	52
<i>Exploratory Data Analysis (EDA)</i>	54
<i>Inference</i>	56
<i>Modeling</i>	56
<i>Spatial Analysis</i>	57
Univariate Statistics	57
<i>Parametric Methods</i>	57
<i>Nonparametric Methods</i>	64
Computer Intensive Methods (CIM)	66
Multivariate Methods	67
<i>Multivariate Analysis of Variance (MANOVA)</i>	67
<i>Multivariate Analysis of Covariance (MANCOVA)</i>	68
<i>Canonical Correlation Analysis (CCA)</i>	68
<i>Principal Component Analysis (PCA)</i>	69
<i>Discriminant Analysis (DA)</i>	69
Interpretation and Presentation of Results	69
<i>Statistical Significance Versus Ecological Significance</i>	70
<i>Relationship Between Statistics and Ecological Risk Assessment</i>	71
4 Summary	73
References	74

Appendix A:	Symbols.....	85
Appendix B:	Glossary of Terms	86
Appendix C:	Acronyms.....	95
Appendix D:	Checklist for Implementation of Field Research for Evaluating Effects of Military Smokes and Obscurants (S/O) on Threatened and Endangered (T&E) Species	96
CERL Distribution		102

1 Introduction

Background

Military smokes and obscurants (S/O) are an integral component of significant training and testing missions. Effects of these compounds on natural habitats and resident populations are not well documented. It is perceived that exposure to these compounds may have adverse effects on threatened and endangered (T&E) species (listed pursuant to the Endangered Species Act, 16 USC 1531-1544) that reside on military installations (Getz et al. 1996). Ecological system responses to natural disturbance regimes and anthropogenic perturbations (disturbances caused by humans) are extremely variable, and often with significant spatial and temporal confounding effects (Noss and Cooperrider 1994). Careful planning and execution of valid experimental designs, sampling strategies, field data collection methods, and statistical analyses protocols are essential for determining the cause-effect relationships among ecosystem elements and chemical agents. Although long-term ecological sustainability of training/testing lands is also an important issue for military readiness, compliance with environmental laws such as the Endangered Species Act is critical to comprehensive land management. Federal listed species exhibit low or seriously declining population densities, or have very limited distributional ranges — often both. Experiments to detect training/testing effects on T&E species populations must, therefore, be very sensitive and possess high statistical power. On the other hand, these experiments dealing with the effects or fate of S/O may by their very nature be relatively expensive to implement and to collect field data compared with typical ecological field studies. Optimal sample size, therefore, must be carefully determined with serious consideration of field and objective-relevant experimental design, statistical power, sampling variance, measurement precision, and valid statistical analysis procedures.

Objectives

The objectives of this phase of the study were to evaluate, select, and recommend sampling strategies and statistical analysis procedures appropriate to determine the environmental effects of military S/O on T&E species and the ecosystems that support them. These sampling designs and analysis procedures will allow

installation natural resource managers and other personnel to use statistically valid techniques to measure effects of S/O on T&E species in order to provide information on which to base actions taken to ensure compliance with the Endangered Species Act and other environmental regulations.

Approach

Current literature on ecological and statistical aspects of experimental design; field sampling and modeling for aerial contaminants; and ecosystem processes relevant to rare plant and animal species was reviewed. Ecological assessment methodologies relevant to identification and characterization of potential direct and indirect S/O effects on T&E species and their habitats were also reviewed. Sampling designs and statistical analysis procedures for assessing the effects of S/O on T&E species and their associated habitats were selected from the synthesis of these reviews and their applications discussed. Many realistic, but hypothetical, interactions between S/O materials and various species have been utilized for illustrative purposes. The authors do not imply, or propose, that such studies are required.

Scope

This report provides a general overview of sampling designs and statistical procedures for assessing the effects of military S/O on T&E species in terrestrial and aquatic ecosystems. A broader technical discussion of ecological design and analysis, which could serve as a companion to this report, can be found in Krzysik (1998a). Laboratory and greenhouse experiments are not addressed in this report. Literally hundreds of designs and analysis procedures are possible; the ones discussed in this report were selected for ecological relevance, simplicity, ease of execution, cost-effectiveness, statistical robustness, and applicability of results with respect to providing Department of Defense (DoD) managers with tools for making effective, defensible decisions on ways to demonstrate the effect or lack of effect of military S/O on T&E species. This report is intended to lay a basic foundation for understanding the following areas as they are applied to studies of S/O effects on T&E species: (1) principles of experimental design and statistical analysis procedures, (2) strengths and weaknesses of each design and analytical procedure, and (3) statistical and ecological rationale for selection of particular designs. Most aspects of this systematic approach are fully applicable to studies of other effects on other species, and are not limited to S/O effects.

This report is designed to assist DoD installation natural resource managers who are concerned with comprehensive land management, compliance with environmental laws, and the long-term ecological sustainability of military training and testing lands. The report can be used for multiple purposes. It is intended to directly assist DoD installation personnel to design field studies that examine possible impacts of S/O on T&E species and to perform related statistical analyses that are within the scope of personnel time commitments and expertise. If more extensive studies are required or the outcome is of great consequence (e.g., military activities subject to severe restrictions by regulators), it is recommended that either a professional statistician be consulted or the study be performed by a contractor with the required expertise. If DoD personnel prepare a Statement of Work (SOW) for work to be performed by a contractor, the information provided in this report can assist in the following ways: (1) information for preparation of statements of requirements included in the SOW, (2) source of guidelines to the contractor regarding how the work is to be performed, and (3) considerations for monitoring contract work.

The exact sampling design and statistical analysis used in a particular situation depends on the specific questions to be addressed. Because extensive field studies (including sampling to realistically measure concentration and duration of exposure) to assess smoke impacts are so complex, consultation with a professional statistician or biometrician prior to conducting the field study is highly recommended. Such an expert can review a study design for statistical validity or recommend appropriate sampling design and statistical procedures for a specific study related to the species that occurs at the specific location under the training/testing scenarios identified by the DoD manager. The problem is not just one of design and analysis, but also of a complex set of logistical and field methods required to actually conduct a realistic study. S/O samples must be related to sampled elements of flora and/or fauna. In other words, biological, ecological, and military training-relevant scenarios, must be tied together in a scientific and experimental context. Sampling methods appropriate for S/O and T&E species are recommended in Volume 2 of this report series (Sample et al. 1997).

Mode of Technology Transfer

The report will be posted to the World Wide Web, making it accessible to installations where S/O and riot-control agents are used and where threatened, endangered, or candidate species are known to occur or may be present. Military organizations particularly concerned with S/O and riot-control agents will be notified when the report is available.

2 Sampling Design Considerations

Introduction

The scientific approach to evaluating S/O effects on T&E species must be based on a rigorous statistical foundation that results in logical planning, design, execution, analysis, interpretation, and presentation of results. Green (1979) describes this sequence as: purpose --> question --> hypothesis --> sampling design --> (experimental execution) --> statistical analysis --> tests of hypothesis --> interpretation and presentation of results. "Experimental execution" was the terminology proposed by Hurlburt (1984). The generalized scheme of a logical research program in Underwood (1997, Figure 2.1) consisted of: observations of patterns in space or time, models of theories or explanations, hypotheses and predictions based on models, null hypothesis as the logical opposite to the hypothesis of interest, experiment as the critical test of the null hypothesis, and interpretation of results. Underwood strongly insists, however, that this is not the end of the research. It is important to further probe the model — both in more generalized and in more specific contexts — with more rigorous testing. This continued research is the recipe for scientific progress and the challenge to established paradigms (Kuhn 1970). Although budget constraints limit sampling scales and replications, it is also important to revisit or extend experimental sites and repeat critical experiments (Connell and Sousa 1983). These approaches work well when inferential statistics (see glossary) and hypothesis tests are appropriate. Some research and issues in the ecological and environmental sciences, however, cannot be resolved with the classical inference approach. Despite the cautions of statisticians, since Berkson's (1942) insightful paper on the use and misapplication of significance tests and hypothesis testing, practitioners of environmental analyses routinely use significance tests as dogma (reviewed in Krzysik 1998a). Also, nonparametric tests are routinely applied to unbalanced, small sample, noisy, heterogeneous, highly skewed, or carelessly collected, "messy" data sets, in the mistaken belief that these tests are robust or independent of statistical assumptions, inadequate sampling, or poor research design (Krzysik 1998a) (see nonparametric section for expanded discussion and references).

Statistical analysis of environmental data is challenging for a number of important reasons: (1) acquiring adequate sample sizes and replicates, (2) sampling

independence in replicates and treatments, (3) unbiased sampling, and (4) parametric assumptions of the data. It is important to consider parametric assumptions of the data: **normality** – the data are distributed as a normal or Gaussian “bell-shaped” distribution; **homoscedasticity** – similar variances exist among all comparison groups; **independence** of sampling errors; **homogeneity** of experimental units – areas or populations sampled have similar characteristics; and **additivity** of error effects for treatments – each treatment affects only the experimental unit to which it is applied, allowing for the detection of true differences between treatments. Lack of true independence in treatment replications (the classical approach) presents the most formidable and hardest to overcome problem for environmental field studies (Hurlburt 1984; Underwood 1997). Importantly, in the case of military S/O research, careful controls on the amount, location, and timing of smoke releases would be necessary for validating the significance among treatments. This represents a significant challenge in coordination with military activities for research designs under actual field conditions. These constraints can present considerable obstacles in a researcher's attempts to characterize S/O effects on T&E species populations and habitat. Eberhardt and Thomas (1991) summarized the challenges faced by ecologists in characterizing natural resources:

Unfortunately, natural systems appear to be very “noisy” in the sense of stochastic (chance) fluctuations, and environmental research techniques are subject to substantial “measurement errors,” i.e., they rarely measure anything exactly and consistently. In such circumstances it seems desirable to adhere to the more flexible viewpoint ... in which a long series of successive studies each yield a “decision” (based on statistical tests), but a “conclusion” (a scientific law, perhaps), ultimately depends on a reassessment of this whole series of individual results. Such an outcome is generally unattainable under the rules of strict logic ...

In circumstances where controlled experiments utilizing inferential statistics (see glossary) cannot be used, observational studies using descriptive statistics (see glossary) may provide ecologically meaningful answers to questions about S/O effects. Most progress in ecological and environmental research has been made by combining the experience and observations of the researcher with controlled experimentation (Eberhardt and Thomas 1991).

An introductory summary to research and experimental design, common pitfalls and problems with experimental designs and statistical analysis, and guidelines for designing ecological or environmental monitoring programs can be found in Krzysik (1998a, 1999). For natural resources and land managers not familiar with statistics and data analysis, a number of excellent introductory books are

available (Kachigan 1986; Campbell 1989; Motulsky 1995; and Zolman 1993). Basic fundamental statistical analysis textbooks that are found in the classroom and in the libraries of working field biologists are: Steel and Torrie (1980), Snedecor and Cochran (1989), Sokal and Rohlf (1995), and Zar (1999). Readers of this report may find Appendices A through C useful for symbols, terms, and acronyms used herein.

Purpose: Classification of Field Research — Mensurative and Manipulative Studies

Introduction

The experimental designs of environmental studies can be broadly classified as mensurative or manipulative (Hurlburt 1984). *Mensurative* studies involve simple observation or measurement of intrinsic ecological phenomena. The researcher makes no attempt to manipulate or influence events (i.e., apply a treatment) during the course of the study; instead, time or space are used as treatment variables, and inherent properties of the populations or systems are the features of interest. *Manipulative* studies, typically using the experimental designs of researchers, are characterized by the application of different treatments to different experimental units (e.g., releasing specific amounts of white phosphorus [WP] smoke into different areas and evaluating the effects). Both inferential and descriptive analysis techniques can be used with mensurative and manipulative studies.

Eberhardt (1976), Hurlburt (1984), Eberhardt and Thomas (1991), and Underwood (1991, 1992, 1994) reviewed the issues of mensurative and manipulative studies and pseudoreplication, bringing renewed attention to the difficulties of achieving true replication in ecological experiments and environmental field settings. The problems encountered in meeting the assumptions and challenges of experimental design principles have been recognized for some time by researchers outside of laboratory settings, and environmental and social experimental designs have been referred to as "quasi-experimental designs" (see discussion and references in Krzysik 1998a). Milliken and Johnson (1989) provide a discussion and practical guidance for the analysis of unreplicated experiments.

To clarify the purpose of a research effort, several items should be considered (Taylor 1990). These items are the desired outcome of the investigation, the population of concern, the parameters of interest, the facts already known about the situation, assumptions needed to initiate the investigation, the basic nature

of the problem, and temporal and spatial aspects of the problem. See Appendix D for a checklist of recommended background information for use in chemical impact studies.

Desired Outcome of the Investigation

The researcher should be able to describe the information he/she needs to obtain or what he/she wishes to demonstrate as a study result. To describe the desired final outcome, the researcher needs to define the specific problem or issue to be resolved and the criteria that will be used to determine if the research goals have been met. Examples of outcomes are (1) to record changes in a population over time, (2) to compare two or more populations with each other, and (3) to document compliance with state or Federal regulations.

Population of Concern

The ecological definition of a population is different from the statistical definition. Ecological populations are spatially, temporally, and genetically coherent groups of plants or animals of a given species or subspecies. In other words, they constitute a group of individuals that are characterized as freely interbreeding, and they are, in theory and under natural conditions, reproductively or spatially separated from other populations (Sutton and Harmon 1973). Statistically, a population is the set of numbers that describe all possible events in a defined universe. If, for example, the defined universe is the liver-tissue concentration of hexachloroethane (HC) compounds in golden-cheeked warblers (*Dendroica chrysoparia*) on a given installation, then the corresponding statistical population of concern is the set of all possible numbers that could describe this concentration. This set of numbers is bounded, continuous, and infinite (e.g., HC concentration may equal 0.005, 70.89, 116.6, 500, or 999.99999 mg per gram of liver tissue, etc.), in contrast to the ecological population of warblers, which is bounded, discrete, and finite (e.g., 220 warblers). A convenient distinction between the two types of populations is that a statistical population is composed of numerical values (but may also correspond to the real population through a frame of reference), while an ecological population is composed of biological entities. Both the ecological and statistical populations of concern should be identified prior to the initiation of the study.

Parameters of Interest

A parameter is a fixed numerical quantity that describes a characteristic of a population (Iman and Conover 1983). Parameters are constants that define location and moments of statistical populations (Winer, Brown, and Michels 1991).

The most useful location parameter is the arithmetic mean, but other examples are geometric mean, median, and mode. Moment parameters define the frequency distribution of the statistical population, and include standard deviation, standard error, variance, skewness, and kurtosis. Parameters are denoted by lower-case Greek characters such as μ , σ , or ρ . In practice, it is rarely possible to measure all individuals in a population, so subsets of the population (samples) are measured. Unlike population parameters, which are constant, numbers used to summarize sample data are variables that change with every sample taken. The numbers that characterize sample data are called statistics, variables, or sample estimators for the population, and are denoted by lower-case English letters (e.g., \bar{x} , s , r), or by Greek letters with "hats" or carets (e.g., $\hat{\mu}$, $\hat{\sigma}$, $\hat{\rho}$). Therefore, "statistics" fundamentally represents the study of the numbers that characterize sample data and how effectively they describe the larger population of interest. For more details, consult a basic textbook in statistics (e.g., Zar 1999).

When conducting cause-effect environmental studies, the researcher should use a consistent and logical process to select parameters of concern based on the objectives and goals of the study. Ideally, it is important to identify parameters that (1) provide the most information, (2) can distinguish between anthropogenic impacts and natural environmental variability, (3) are reliable and sensitive indicators of change, (4) are the most cost-effective to measure, and (5) possess additional important interests or merits (Green 1979; Landis et al. 1994 [specific for risk assessment]). Usually, the researcher must choose among these criteria to optimize some desired attributes at the expense of others. Statistical methods such as linear regression or discriminant analysis could be performed on preliminary data sets to possibly identify diagnostic variables that may be able to distinguish between natural and anthropogenic effects. Identification of diagnostic variables by inspection of graphed data is another technique that could be used.

Facts Already Known About the Situation or Problem

Basic information about the problem should be collected in a systematic manner and evaluated. Such information may include:

1. listings of potential and field-verified T&E species populations on the installation
2. maps of T&E species habitat
3. locations of T&E species sightings or maps of population distributions
4. identification of critical habitat needs for T&E species (e.g., habitat extent, successional stage, food/water/nesting/shelter resources)
5. life history of T&E species
6. past and current T&E species population trends

7. ranking of research priorities based on military activities most restricted by T&E species, T&E species population trends, future anticipated use of the area, etc.
8. timing, frequency, intensity, duration, and location of military exercises using S/O
9. delineation of areas where T&E species and training activities coincide
10. types and quantities of S/O released
11. prevailing weather conditions, wind direction, topography, S/O dispersion patterns
12. known physiological or behavioral changes caused by exposure to S/O (e.g., bioassay results)
13. types of nonmilitary chemicals released on or near the installation (e.g., herbicides, insecticides, fungicides, fertilizers, output from manufacturing plants)
14. land use and ecological history of the area where S/O exercises occur
15. the nature of any regulatory constraints on military activities
16. labor and financial resources available to address the issues.

The nature, amount, and quality of preliminary information available for evaluation directly affects the decisions made with regard to the type of study to conduct. Information may be obtained from personal observation of the situation that needs to be addressed, research results for similar studies, literature reviews, or expert consultation. Conducting a pilot study is a very valuable way to collect some preliminary data that may reveal new aspects or problems that were not previously identified. Talking to resource managers who deal with some aspect of these issues on a regular basis may provide additional insight from a different viewpoint. Of particular value would be communication with other state and Federal land management agencies (e.g., National Park Service, Bureau of Land Management, Forest Service), but also regulatory agencies (e.g., U.S. Fish and Wildlife Service, Federal and state environmental protection agencies), research organizations, and universities that may also provide information to help clearly define the purpose of the study.

Assumptions Needed To Initiate the Investigation

Once initial information has been collected, the researcher should identify and describe any assumptions or constraints that affect the study. Such assumptions are not necessarily easy to formulate and require careful thought. Statistical assumptions which should be delineated include the distribution of the data, the presence or absence of spatial, temporal, or other patterns in the data, the estimated effects of military or nonmilitary activities that might affect data interpretation, and limitations of the sampling design and methods. Ecological assumptions include an initial estimate of the nature and extent of the problem to be studied, the species and specific populations likely to be affected by S/O, and

the informed estimates needed to initially replace knowledge gaps about the species/populations and S/O under investigation.

Basic Nature of the Problem: Research, Inventory, Monitoring, or Conformance

Research studies typically have a very narrow focus and attempt to answer one or a few highly specific questions. Replication of experimental units may be required to achieve an estimate of experimental error, and sample sizes that are adequate to achieve desired power or precision may be needed. Single inventory or assessment studies provide a "snapshot" of population status or characteristics. Monitoring studies are conducted to evaluate the nature and extent of changes in the population over a period of time or how populations vary spatially with time. Conformance studies, which are conducted to demonstrate an installation's compliance with environmental regulations, may incorporate elements of research, inventory, or monitoring studies, but their main purpose is to show that specific legal requirements are being addressed.

Temporal Nature of the Problem: One-time, Short-range, or Long-range

Different aspects of the research problem being evaluated may span several different time scales. For example, the physical presence of a single S/O release in an ecosystem may last for minutes, but the long-term effects of repeated S/O releases in the system may require decades to detect. Ecological time scales may encompass a single life stage (e.g., larval stage), the lifetime of an organism, or long-term succession of a plant community (U.S. Environmental Protection Agency [EPA] 1992). Each type of problem requires a different approach with respect to the number of times samples will be taken and the time intervals between sampling efforts. Complex research efforts and sampling designs may require the coordination of data collection across a range of temporal scales.

Spatial Nature of the Problem: Local, Regional, or Global

The area affected by S/O releases may range from highly localized sites for smoke grenades, to hundreds of hectares when large-scale military maneuvers are conducted. Smokes also disperse and become a component of the global atmosphere. Some areas may be limited in size or have unique features that are not found elsewhere. Replication may not be possible in such areas. The distribution of T&E species populations and their habitats across a landscape also directly determines the nature of the sampling process.

Question: Definition of Objectives

The previous section discussed considerations for field studies of T&E species and S/O in general. For a specific study, the objectives of the field study should be clearly defined in advance. The objectives should be focused and specific, quantifiable, verifiable, and relevant to the needs of the particular installation. Each variable selected for measurement should directly contribute to achieving the objectives. Korte, Klein, and Sheehan (1985) identified three aspects of recognizing and predicting environmental hazards: (1) the detection, quantification, and prediction of the environmental behavior of chemicals, (2) the diagnosis of toxic effects and the estimation of their magnitude, and (3) the estimation of exposure. Another important consideration is relevance to structure, function, or processes of ecological elements. Examples of general objectives and more specific applications that can be tailored to meet individual research requirements are given in the numbered list following this paragraph. Comparisons across time (objective 4) and location (objective 6) can be included as part of the objectives. Objectives 1 and 2 focus on the presence or absence of chemical effects on T&E species populations or habitats, objectives 3 and 4 focus on the magnitude and duration of these chemical effects, and objectives 5 through 8 focus on the concentration levels of chemicals in biotic or abiotic systems without making predictions about their biological effects. Objectives 7 and 8 are different because chemicals that do not bioaccumulate would not necessarily be present in body tissues or organs, but may still have an effect on an organism. Objective 9 goes beyond simple direct effects to consider combined effects, either simultaneously (interactions) or over time. Note: the examples given are realistic examples, but do not represent current or proposed research.

1. To determine if smoke usage results in adverse biological effects for T&E species.

(Note: In most cases T&E species surrogates would be used to make the determinations.) Primary biological effects include: survivorship, reproduction, physiological changes, and behavioral abnormalities.

Examples:

- a. To determine if HC smoke exposure results in decreased photosynthesis for the hooded pitcher plant (*Sarracenia minor*).
- b. To determine if fog oil exposure results in decreased hatchability for eggs of red-cockaded woodpeckers (*Picoides borealis*).

2. To determine if T&E species habitats and environments are adversely affected by military smokes.

Examples:

- a. To determine if food sources (prey base) for the Mexican wolf (*Canis lupus baileyi*) are declining as a result of exposure to WP smoke.

- b. To determine if the oxygen content of streams inhabited by bluestripe shiners (*Cyprinella callitaenia*) is declining as a result of contamination by fog oil.
3. To determine type, magnitude, and duration of S/O effects on T&E species (using surrogate species).

Examples:

- a. To determine number, size, and distribution of pulmonary lesions caused by graphite flake inhalation for St. Andrew's beach mouse (*Peromyscus polionotus peninsularis*).
- b. To determine the level of foliar injury in relict trillium (*Trillium reliquum*) caused by root uptake of nickel-coated graphite.
4. To determine changes in T&E species populations over time as a result of smoke effects.

Examples:

- a. To determine the rate of change in anhinga (*Anhinga anhinga*) populations as a result of exposure to WP.
- b. To determine the rate of change in Indiana bat (*Myotis sodalis*) populations as a result of exposure to HC smoke.
5. To determine the chemical concentration levels of S/O for soil and water environmental compartments.

Examples:

- a. To determine the total soil chemical load for all smokes released in the S/O training area.
- b. To determine water transport and storage in sediments, and compare lentic (standing waters) and lotic (running waters) environments.
- c. To determine the amount of WP in the aquatic sediments of a wetland area.
6. To compare the accumulation patterns of chemicals with respect to different atmospheric transport processes, topographic features, or ecosystem structures.

Examples:

- a. To compare accumulation patterns of fog oil along riparian zones (on or near bodies of water, esp. rivers) with patterns in nonriparian areas.
- b. To compare accumulation patterns of red phosphorus in trees with accumulation patterns in lichen.
7. To correlate levels of ambient or environmental contamination with levels of tissue and organ bioaccumulation (i.e., total body chemical burden) for S/O in selected plant and animal species.

Examples:

- a. To correlate soil concentrations of HC with HC concentrations in Southern milkweed (*Asclepias viridula*) vascular tissue.
- b. To correlate ambient concentrations of fog oil with fog oil concentrations in golden-cheeked warbler (*Dendroica chrysoparia*) livers.

8. To correlate levels of ambient or environmental smoke exposure with actual dosage intakes of chemical by inhalation, ingestion, absorption, adsorption, or other mechanisms.

Examples:

- a. To correlate ambient WP concentrations with blood WP concentrations of Cumberland pocket gophers (*Geomys cumberlandis*).
- b. To correlate aquatic HC concentrations with dermal HC absorption by shortnose sturgeon (*Acipenser brevirostrum*).
9. To address cumulative effects and interactions such as additive effects, but may also include synergistic (more than additive) and antagonistic (less than additive) effects.

Examples:

- a. To compare the level of foliar injury in relict trillium (*T. reliquum*) caused by root uptake of nickel-coated graphite in an area where there has been limited release of the nickel-coated graphite to the level of injury in an area that has been subjected to repeated releases.
- b. To compare the rate of change in anhinga (*A. anhinga*) populations as a result of exposure to HC smoke with the rate of change as a result of exposure to both HC smoke and fog oil smoke.

Hypothesis: Selection of Correct Conceptual, Estimation, and Predictive Models for Smoke Effects on T&E Species

Definition

Broadly speaking, a hypothesis is a statement of an assumed condition that can be confirmed or refuted by additional testing or observation. Technically, hypotheses can only be rejected; they cannot be proven. The only other alternative is failing to reject a posed hypothesis. Restating the research objective as a hypothesis will more narrowly define the exact scope and thrust of the research effort so that studies can be focused on obtaining the specific information needed to answer the questions of interest. Either a qualitative or quantitative statement of the expected relationships to be investigated may be used as a hypothesis. If inferential analysis is desired, the objective needs to be restated in a manner that can be confirmed or refuted with a known level of confidence by using null and alternative hypotheses. The null hypothesis (H_0) is a formal statement or conjecture to be tested. It is often worded in a way to indicate that no change has occurred or no difference exists (Iman and Conover 1983). The alternative hypothesis (H_A) is a statement that indicates the condition expected to be true if the null hypothesis is rejected. Inferential statistical procedures require the *a priori* assignment of rejection criteria for the null hypothesis (see section

on Data Quality below). Examples of restating an objective as hypotheses for descriptive and inferential analyses are given below.

Objective: To determine if HC smoke exposure results in smaller size for the Florida willow (*Salix floridana*, or surrogates).

Qualitative hypotheses:

- (1) The trunk diameter of Florida willow exposed to HC smoke is less than the trunk diameter of Florida willow not exposed to HC smoke.
- (2) The biomass of Florida willow exposed to HC smoke is less than the biomass of mature Florida willow not exposed to HC smoke.

Quantitative hypotheses:

- (1) The trunk diameter of Florida willow exposed to HC smoke is less than 12 cm. (The researcher already knows from previous studies of the literature that the average trunk diameter for Florida willow in unaffected areas is 12 cm.)
- (2) The biomass of Florida willow exposed to HC smoke is 60 kg/ha less than the biomass of Florida willow not exposed to HC smoke.
- (3) The canopy spread of Florida willow exposed to HC smoke is 20 percent less than the canopy spread of Florida willow not exposed to HC smoke.

Null (H_0) and alternative (H_A) hypotheses:

- (1) H_0 : The trunk diameter of Florida willows exposed to HC smoke is greater than or equal to 12 cm.

H_A : The trunk diameters of Florida willow exposed to HC smoke are less than 12 cm.

- (2) H_0 : The canopy spread of Florida willow exposed to HC smoke is equal to the canopy spread of Florida willow not exposed to HC smoke.

H_A : The canopy spread of Florida willow exposed to HC smoke is not equal to the canopy spread of Florida willow not exposed to HC smoke.

For long-term or complex ecological research projects, developing and testing conceptual, estimation, and predictive models may be necessary to formulate multiple hypotheses and to determine their relative importance in the context of the larger study. Models can also be used to clarify uncertainties in relationships between ecological and chemical entities and to demonstrate possible interactions among various elements in the system.

Conceptual Models

Conceptual models show relationships between chemical compounds and T&E species populations or habitats. They provide the basis for identifying likely interactions between military S/O concentrations in the air or ecosystem and behavioral or physiological changes in T&E species as a result of exposure. More extensive discussion of conceptual models and the current state of the art can be found in U.S. EPA (1998) and Suter (1999). Since the correct selection of conceptual models leads to the selection of appropriate variables and hypothesis tests for statistical models, extensive knowledge of the smoke usage/distribution on the installation and physiological responses and population dynamics for the T&E species populations of interest are necessary.

Both chemical behavior in the environment and organism responses should be included in the formulation of conceptual models. Chemical considerations include fate and transport mechanisms, which can be evaluated by modeling or estimating (1) ambient chemical concentrations under different training scenarios, (2) transformation/decomposition products under selected atmospheric and environmental conditions, (3) deposition and leaching rates in specific ecosystems or strata (e.g., midgrass prairie vs. oak/juniper woodland; understory vs. overstory strata), (4) rates of chemical accumulation or decomposition in soil and water compartments of the ecosystem, and (5) relationships between environmental exposure and actual dosage rates. Environmental considerations for conceptual modeling are (1) adsorption or absorption and ingestion/inhalation pathways for terrestrial and aquatic plants and animals, (2) trophic bioconcentration for target compounds and organisms, (3) transformation, decomposition, and excretory pathways for chemicals in living systems, and (4) predicted physiological and behavioral responses of T&E species to known chemical dosages. Estimates for missing information in conceptual models may come from studies on related compounds in technical literature, personal experience, or expert opinion. See Appendix D for a checklist of recommended background information for use in chemical impact studies.

Estimation Models

Estimation models are sets of mathematical equations that represent the system of interest. They are used to identify variables that contribute to explaining chemical or biological processes and to provide probability estimates for events that affect the system.

Estimation models can be deterministic or stochastic. *Deterministic models* assume that conditions in the equations remain fixed and constant (i.e., no statistical or environmental uncertainty is included in the model), and may be used to describe parameters associated with basic environmental/T&E species states and processes, such as age structure, population size and growth, reproduction rates, and environmental conditions. For example, population growth over time may be calculated with a deterministic model by using a constant growth rate factor multiplied by the population size at a given time. *Stochastic models* are used to introduce random fluctuations in the system. A population growth model that incorporates stochasticity may use the basic deterministic model modified by the inclusion of probabilities for chance events such as famine, drought, predation, or chemical impacts.

When building estimation models for determining effects of S/O on T&E species populations or habitat, the researcher should consider both factors that affect short-term population fluctuations (i.e., variability) and factors that affect long-term abundances (i.e., means). For example, a T&E species population that experiences a drastic decline in one year may not be able to recover, even if the mean population size appears to be increasing on the basis of long-term trends. On the other hand, a comparatively stable population with minor fluctuations in size may not survive if it is experiencing a gradual, but persistent, long-term population decline (Burgman, Ferson, and Akcakaya 1993). Information gained from estimation models can be used in population viability models or risk assessment models to allow natural resource managers to determine the probabilities of having unacceptable conditions (e.g., T&E species population levels below a critical recovery point) or for identifying the likelihood of occurrence for worst-case scenarios given different mixes of environmental conditions and military activities. Selection of appropriate estimation models requires the identification of variables to be evaluated and their relationships to each other, assignment of probabilities to random events, and selection of appropriate statistical tests to measure effects.

Predictive Models

Estimation models are used to test inferences within the temporal and spatial boundaries of the data collected. As the models are tested with data collected from the field, some variables may be found to have important effects on the system of interest, while others may have little or no effect. As more data are gathered, the estimation models can be refined into predictive models, which are used to characterize system behavior beyond the range of the data (Zar 1999). Development and interpretation of S/O predictive models should be done with caution, because S/O behavior is so variable and in many cases it may be very difficult or impossible to extrapolate extended effects in time or space from available data. Predictive models are more likely to be successful when the researcher can control sources of natural variability and sampling error. For example, models that forecast atmospheric dispersion of S/O may be more difficult to validate than models that predict S/O effects on soil microorganisms, because of differences between the two types of studies with respect to sampling ease, repeatability, availability of monitoring equipment, timing and logistical constraints, inherent system variability, and other considerations.

Sampling Design: Development of Appropriate Strategies for Allocating Treatments and Collecting Samples

The experimental or sampling design, in simplest terminology, is the set of plans and instructions by which the data are collected and specific statistical design protocols are met (Green 1979; Iman and Conover 1983; Underwood 1997). The experimental design can be mensurative or manipulative. The difference between the two terms relates to whether the researcher will intervene (or apply treatments) in the study, or whether he will simply observe events as they happen without attempting to control or manipulate them. Developing a good experimental or sampling design requires the determination of the true population to be sampled; selection of appropriate variables to measure, experimental units, and sampling units; awareness of special considerations for sampling biotic and abiotic media; identification of confounding factors; and scale considerations. Criteria such as replication and independence must be applied (Hurlbert 1984). If the experiment is a manipulative experiment, the kind and number of treatments to be applied should also be specified, and the details of assigning treatments to experimental units explained.

Determination of True Population To Be Sampled

In statistical terms, a population is the set of all possible values of a variable (Steel and Torrie 1980). When every value of a population is known, then the population is completely defined. A statistical population can be large or small, finite or infinite, and can consist of discrete or continuous numbers. An example of an infinite population consisting of continuous numbers would be all possible values for the chemical concentration of fog oil taken at a height of 10 m and a distance of 50 m from an M3A4 generator after the generator has been running for 60 minutes. An example of a small statistical population consisting of discrete numbers would be the number of young successfully raised by a specific pair of red-cockaded woodpeckers (*P. borealis*) during a 5-year period. In risk assessment analysis, the assessment or measurement endpoints constitute the statistical population to be sampled. The correct descriptions of the statistical and ecological populations to be tested are important in determining the statistical analysis procedures that will be relevant for the study and the extrapolation of the results to a larger context.

Two common mistakes that researchers make when defining the population of interest are (1) inadequate or incorrect definition of the population of interest, and (2) defining one population but sampling a different population or a subpopulation (Green 1979). For example, a researcher may define the statistical population of interest as the ground deposition levels for WP on two training areas of an installation. Such a definition does not take into account factors such as accumulation of phosphorus over time, transformation of WP into other compounds, or the chemical instability of phosphorus compounds under various temperature and humidity regimes. Since phosphorus levels exhibit temporal variability, a better definition of the population would include a restricted time frame and season.

It is also important to adequately define the ecological population to be sampled. An ecological population is a group of genetically compatible individuals with the spatial and temporal potential for reproduction (i.e., a gene pool). Sampling the wrong ecological population can occur when incorrect assumptions are made concerning population distribution and density parameters, home range, or dispersal behavior and parameters. Sampling can also be inadequate or biased. Often, some field sampling strategies may collect only a biased subpopulation of the intended target population (Green 1979). Common causes of subpopulation bias include capturing slower, older, or diseased individuals; larger individuals that are more easily seen or susceptible to being caught in a wider range of net mesh sizes; or brightly colored or strikingly patterned individuals that are more visible. Other causes are behavioral differences in age or gender classes and the

phenomenon of "trap-happy" or "trap-shy" species or individuals. Trapping rodents during the breeding season may capture a disproportionate number of males, because the females spend more time in the nest with their young. Similarly, females in many species of lizards and salamanders may be clutching eggs in subterranean microhabitats. Devices used to capture aquatic species in a lake may collect only slow-moving fish or fish that congregate at a particular depth, which may not be representative of the lake population as a whole. Researchers should be aware of the limitations of their sampling devices and methods in order not to extrapolate data beyond justifiable limits. Borgman and Quimby (1988) defined three populations that must be considered when developing sampling plans: (1) the target population, (2) the accessible population, and (3) the actually sampled population. The study should be designed so that the population actually sampled is representative of the target population. This consideration is especially important if tracer compounds are used to mimic S/O behavior or if surrogate species are used to estimate the effects of S/O on T&E species populations (see **Interpretation and Presentation of Results** in Chapter 3).

Special Considerations for Sampling Abiotic and Biotic Media

Obtaining samples that adequately characterize the actual condition of a system is a formidable task. Variability in the samples, in the size and distribution of the sampled population, and in the sampling and analytical methods all contribute to the uncertainty of the final results. In fact, accurate quantification of the uncertainties associated with sample selection may not be possible. In such instances, the researcher needs to take special care to report qualitative descriptions of the factors that affect sampling accuracy, and any underlying assumptions in the sampling design that affect interpretation of results.

Air.

Military S/O usually consist of exotic materials and properties not found in industrial and agricultural air pollutants. Because of these differences, conventional pollution dispersion models, sampling and analytical methods, and field research techniques may not always be applicable, and new or re-parameterized models and methods need to be developed (Liljegren et al. 1989; Policastro et al. 1991). S/O that contain irregularly shaped flakes or fibers have dispersion characteristics very different from the spherical particles commonly found in industrial pollutants (Bowers and White 1992). The release modes for military S/O also differ from standard industrial and agricultural practices. Industrial releases into air are usually from tall stacks at a single location, but may travel several miles. Agricultural releases into air from aircraft or ground-based equipment typically are spread over several to many hectares. Military smokes

are released at or near ground level over a relatively small area, but may rise and disperse in a plume for a few kilometers or, in the case of a signal smoke, for example, may remain in a comparatively small area (e.g., a fraction of square kilometer). Some S/O materials (e.g., WP) are fired by artillery. Models developed specifically to deal with military smokes have been extensively researched at Dugway Proving Ground, a U.S. Army Test and Evaluation Command (ATEC) test center for smokes and obscurants (Bowers and White 1992). Such models should be used when possible to predict military smoke behavior instead of standard EPA regulatory models.

Studies by Farmer and Davis (1986), Liljegren et al. (1988), and Haines (1993a, 1993b) have indicated that ambient S/O concentrations more than 100 m from stationary release points may be too low for sampling instruments to register. Even samples taken within the 100-m boundary may not be distinguishable from background concentrations of the chemicals of interest. In addition, samples taken within the 100-m boundary may be unreliable because of atmospheric disturbances (especially sudden shifts in wind direction), cross-contamination, loss of volatile sample material, or logistical problems in handling samples. Concentrations of S/O released from moving vehicles tend to be even lower than those released from stationary points for two reasons: (1) greater initial smoke dilution and (2) spread of S/O over a larger area (Bowers and White 1992). Preliminary sampling is highly recommended in order to calibrate sampling instruments, determine the range of S/O concentration to be detected, and avoid wasted sampling efforts. Liljegren et al. (1988) failed to collect any fog oil concentration data in 8 out of 11 experiments because their collection devices were spaced at 100-m intervals up to 1,600 m, but valid observations for fog oil could only be detected within 25 to 75 m of the release site (i.e., the resolution of the sampling design grid was too coarse to capture the fog oil released).

Extremely sensitive instruments with specialized calibration and operation requirements are typically necessary to quantify ambient chemical concentrations. Air sampling is difficult even with highly trained personnel and specialized equipment. Chemical agents may be released as aerosols, volatilized liquid droplets, particulate matter, or mixtures; each phase requires different sampling instruments and techniques. Keith (1991), Haines (1993b), Liljegren et al. (1988), and Farmer and Davis (1986) provide excellent suggestions for recommended sampling considerations and sample preservation strategies for various chemical mixtures, while Nam et al. (1999) provides information for specific smoke, obscurant, and riot-control agent chemicals.

The behavior of S/O is highly dependent on the weather conditions present during their release. Therefore, results from one study should not be generalized

across broad ranges of weather conditions. Bowers and White (1992) described the lifetime of a fog oil droplet as ranging from 0.22 seconds to over 39 years for ambient temperatures of 40 °C and -30 °C, respectively. Photoreactivity and the presence of other reactive chemicals also affect the length of residency for obscurant compounds suspended in the air. Wind speed, atmospheric stability, humidity, and precipitation play important roles in the mixing and dispersion of volatile compounds (Keith 1991; Farmer and Davis 1986). Haines (1993a) summarized the variability of fog oil, aluminum, brass, phosphorus, and other S/O concentrations in ambient air as follows:

Air is also an extremely variable medium in which concentrations of materials can vary naturally by orders of magnitude due to changes in the on-site meteorology and localized contamination. Because of this variability, air is a recalcitrant sampling medium. Results from air sampling at the same location but at different times of the day can differ by orders of magnitude due to changes in predominant wind direction and on-site activities. Because of air's variability, all but the most severe analytical errors will be overwhelmed by errors in extrapolating the data from a limited period to a much longer period and from a limited area to a much larger area. Therefore, many statistical methods that are used to assess data from other sources are not applicable to air data.

Some laboratory analysis procedures for certain S/O may present special difficulties (Haines 1993a). Fog oil concentrations are often measured using the Total Recoverable Oil and Grease method (TROG, EPA Method 413.2). This technique, although considered one of the best general oil analysis methods currently available for assessing fog oil concentration, has several serious disadvantages. The method has been found to be unreliable, and vigorous efforts to find a better process are being pursued (Haines 1993a). Considerable variations in results are possible as a consequence of procedural differences allowed by the method; therefore, laboratory protocols must be strictly delineated in advance. In addition, since TROG measures total oil, it cannot distinguish between obscurant hydrocarbons and other hydrocarbons (e.g., diesel fuel or agricultural chemicals). Haines (1993a) also noted that sample weight made a difference in the concentration of fog oil recovered. Less fog oil was generally retrieved from larger samples in a controlled study (i.e., fog oil concentration was diluted in the larger samples). This dilution problem needs to be addressed for all time-series S/O studies because, if different sample sizes are compared, the results may be invalid.

Biota.

The body size, trophic position, and developmental stage of organisms are all important factors on the effects of chemicals, including S/O (Kendall and Lacher 1994), and all three should be evaluated. These factors are also important considerations for sampling body tissues or fluids for S/O concentrations. Chemical tissue concentrations may accumulate in organisms, and some chemicals even biomagnify in species that are high in the food chain (i.e., top predators) (Di Giulio et al. 1995). The potential for accumulation or biomagnification varies with S/O material, but should be considered. Also possessing a higher potential for biomagnification are larger organisms within a given trophic level, as they are usually higher in the food chain as well. Organisms in embryonic or early developmental stages may be especially sensitive to chemical stressors because of the high cellular activity and metabolism of rapidly differentiating and multiplying cells. The combined effects of higher S/O tissue concentration with the more rapid and variable growth patterns of young organisms may increase the effects of S/O significantly beyond what would be expected for an identical concentration in an adult.

The developmental instability (D.I.) approach and technologies (Graham, Freeman, and Emlen 1993) may be useful for assessing or monitoring the effects of S/O on target populations or in ecological communities. The response of organisms to stress is the basis of environmental adaptation, natural selection, and evolutionary potential. D.I. is a powerful and sensitive test system to quantify stress response of individual organisms and has been effectively used with a broad variety of stressors including air and water pollution, grazing, heavy metals, organic toxicants, excess nutrients, temperature, etc., in a wide variety of terrestrial, fresh-water, and marine ecosystems (reviewed in Møller and Swaddle 1997). Animals (Freeman, Graham, and Emlen 1994), plants (Alados et al. 1998), and algae (Tracy et al. 1995) have all been successfully used for analysis. When developing organisms are exposed to stressors, developmental homeostasis is compromised and further growth patterns may become asymmetrical (Freeman, Graham, and Emlen 1994). D.I. is usually estimated as the variance in a trait repeated within the individual, and involves some aspect of symmetry (Graham, Freeman, and Emlen 1993; Freeman, Graham, and Emlen 1994). Random deviations from all types of symmetry have been used as indicators of stress. Unless there is some predisposition for traits to exhibit a certain handedness, the two sides should be mirror images of each other (i.e., they should exhibit bilateral symmetry). The most common measure of D.I. is fluctuating asymmetry based upon the absolute value of the difference in the value of a trait measured on the right and left sides of the body (Palmer and Strobeck 1986;

Graham, Freeman, and Emlen 1993). Fluctuating asymmetry was associated, for example, with a fruit fly species as it declined to extinction (Tsubaki 1998).

Preserving tissue or fluid samples for laboratory analysis requires advance planning to ensure that sample integrity is maintained throughout the collection, transport, and analysis process. Guidance for samples containing selected S/O materials can be found in Nam et al. 1999. Simini (1992) provided some excellent suggestions for field sampling protocols to follow when collecting, handling, and preserving vegetation for the analysis of chemical warfare agents. Such protocols could be adapted to S/O compounds in order to maintain high quality samples for analysis.

Soil.

Special sampling and handling techniques for soil samples are necessary to avoid or minimize: loss of volatile compounds, oxidation-reduction reactions, or transformations by microorganisms and other biological activity. The cost of collecting additional field samples (once in the field) is sometimes inexpensive relative to the cost of getting to the collection site and the cost of laboratory analysis. It may be desirable, therefore, to collect supplementary or redundant samples. Cross-contamination should be avoided by thoroughly cleaning equipment between each sample, and chemical interactions between soil samples and sampling devices should be avoided by using samplers constructed of the appropriate materials. Keith (1991) recommended stainless steel collection devices for soils contaminated with organic compounds, and high-density polyethylene devices for soils contaminated with inorganic compounds. Sandusky (1992) outlined field sampling protocols to follow when collecting, handling, and preserving soils contaminated with chemical agents such as nerve gas or other compounds used in chemical warfare. These protocols could be modified for the S/O under consideration to maintain soil sample integrity for laboratory analysis.

A common problem with collecting representative soil samples in a time-series design is that military activities or burrowing animals may mix contaminated and uncontaminated soils within the soil matrix. Leaching as a result of flooding events may move some S/O compounds into a lower soil horizon, while underground migration of chemicals to or from adjacent areas may create unexpected pockets of lower or higher concentrations. Rising and falling water tables may also affect contaminant levels. The researcher should study the site and soil characteristics of the study area carefully to determine if confounding influences may be present.

Water.

Obtaining representative water samples is very difficult because of the spatial and temporal heterogeneity of aquatic systems (Keith 1991). An important consideration is that biota in lotic (running water) ecosystems may be impacted by a single short-term release (spill) or "pulse" of a toxic compound (MacKay, Burns, and Rand 1995). Subsequent chemical analysis of water chemistry would not reveal the nature of the impact event. Depending on the toxicant, its concentration, and length of exposure, benthic (e.g., from a lake bed or river bottom) cores may be able to detect its impact.

The behavior of a chemical compound in water depends on several factors, the most important of which are: (1) the solubility of the compound, (2) the temperature of the water, (3) the specific gravity of the compound, (4) the nature of the aquatic environment (e.g., lotic or lentic systems, marine, brackish, or freshwater ecosystems, water chemistry), and (5) the size and depth of the water compartment (e.g., ditch, small pond, river, large lake, or ocean). A problem commonly encountered with larger bodies of water is that various chemical compounds and aquatic species may be stratified at different depths. Thermal stratification of water can also complicate the sampling process, as chemicals may exhibit different reactivities at different depths depending on water temperature and redox (oxidation reduction) potential. Flowing water presents special challenges for sampling because mixing within the water column introduces high heterogeneity into the sample.

Keith (1991) recommended that the length of a sampling study for a body of water be approximately 10 times longer than the period of interest in order to effectively characterize the extent of the heterogeneity present. Keith also warned that water sample contamination is a continual problem, which increases in importance as analyte concentration levels decrease. Since water samples are in a continuously dynamic state, their composition may be substantially altered between collection and analysis by chemical, biological, or physical processes. As discussed earlier, toxicological analysis of lotic ecosystems is difficult to assess.

Identification of Confounding Factors

Confounding factors are influences other than the ones being explicitly studied which affect the response of a system. The researcher must consider how to deal with confounding factors when designing the study to reduce the probability of obtaining spurious results and to demonstrate that only the factors of interest contributed to the effects observed. Taylor, Johnson, and Anderson (1994) noted that deriving ecologically meaningful trends of pollutant effects over the long term may be very elusive. When the intrinsic variability of S/O use and other military training impacts are added to the natural variability of ecosystems, separation of effects becomes especially difficult. Some of the more common confounding factors for general atmospheric pollutant studies are as follows (Taylor, Johnson, and Anderson 1994; Winner 1994):

1. Seasonal and diurnal fluctuations of ambient chemical concentrations due to light, temperature, humidity, and wind conditions.
2. Compensatory growth by organisms to offset damage caused by air pollution.
3. Multiple natural and anthropogenic stressors in the environment (e.g., lack of water, light, nutrients, or presence of nonsmoke pollutants), including naturally occurring organics.
4. Secondary response mechanisms (e.g., organisms may exhibit compensatory growth to counteract air pollution damage, but then outgrow their resource base or lower their tolerance to other stressors as a result).
5. Differences in individual and species-specific responses to the same level of chemical stress.
6. Interrelationships between spatial distribution patterns, concentration levels, and exposure time of chemicals in sensitive ecosystems.
7. Presence of both positive and negative S/O effects. S/O may enhance growth and physiological functions for some organisms. For example, fog oil may provide carbon as a food source for certain microorganisms. In turn, the enhanced microbial populations may accelerate secondary succession. Phosphorus, nitrates, potassium, and iron are major plant nutrients. These elements and others that are micronutrients will benefit plant growth, and may coincide with negative effects for other organisms (e.g., pellets of phosphorus may be deadly to waterfowl when ingested [Racine et al. 1992]).
8. Indirect effects that reduce competitive ability, nutrient use efficiency, or other behavioral or physiological responses.
9. Organisms, especially at lower trophic levels and over long time spans, may adjust to the presence of S/O by physiological, behavioral, or genetic adaptations.
10. Organisms may not respond to chemical stressors except at specific times or under specific conditions when they are sensitized to the stressor (e.g., during gestation, molting, or budbreak; during extended drought; during larval stage).

11. S/O may affect reproductive fitness. The effects could be very small and subtle, and difficult to measure or quantify. Reproductive fitness is probably the most important biological factor to monitor, because of its direct relationship to population viability. Even under ideal environmental conditions and circumstances, however, reproductive fitness is very difficult to assess and monitor.

Important confounding factors present in military S/O exercises that may need to be considered are physical habitat disturbance and noise created by tactical vehicles and personnel. Additionally, the use of S/O over a period of years presents the potential for environmental accumulation of persistent materials (Passivirta 1991). Depending upon the S/O material under investigation, persistence may need to be considered.

Some chemicals (e.g., certain pesticides) are known to persist in the environment for as long as several years (Kendall and Lacher 1994; Brown 1978). For S/O, persistence resulting in effects may be more likely for some older smoke materials such as HC smoke (Shinn, Sharmer, and Novo 1987), which is no longer manufactured in the United States, and brass (Wentsel 1986). Graphite flakes, a replacement for brass, are persistent in the environment, but few effects have been documented (Guelta and Checkai 1995). Some components of fog oil, at least prior to the 1986 military specification change (MIL-F-12070C), had the potential to accumulate (Bausum and Taylor 1986). Analyses at two sites where fog oil had been released for several years, however, failed to identify hydrocarbon residues that could be traced specifically to fog oil (Brubaker, Rosenblatt, and Snyder 1992; 3D/International Inc. 1996).

In addition to the persistence of S/O in the environment, cumulative effects should also be considered. Cumulative effects may be important (Riha 1988), and interactive effects may be as well (e.g., synergisms) (Jernelov, Beijer, and Soderlund 1978), but they are likely to be unknown and unappreciated.

Selection of Appropriate Variables

Numerous combinations of responses, ecosystem components, and organizational levels of ecological populations can be evaluated to assess the effects of S/O on T&E species populations and habitats. Relevant examples include: bioaccumulation or bioconcentration of chemicals in tissues and organs (Landrum, Harkey, and Kukkonen 1996), physiological changes in cells or tissues, changes in genetic structure, physical or behavioral changes in individual organisms, population dynamics of individual species, competitive or mutualistic interactions and changes between animal species, successional pathways for plant communities,

and basic process changes in ecosystems. Measurable physical properties of S/O include: concentration, deposition rates, mass extinction rates, and particle sizes.

Selection of relevant variables by the researcher depends on many important factors: ecological relevance, level of sensitivity with respect to the change to be detected, ease or difficulty of obtaining representative samples, and contribution of each variable to the goals of the study. In addition, if the research is being conducted to satisfy environmental compliance requirements, the variables selected for evaluation must meet additional criteria with respect to satisfying policy goals and societal values (Landis et al. 1994).

When studies are conducted over periods of months or years, researchers must be cognizant that ecosystems are spatially and temporally dynamic. Succession and natural disturbance regimes, not to mention inherent environmental variability, will always be factors continually and usually unpredictably influencing measurement and variance of variables. Design criteria and statistical techniques are necessary for accounting for background variability, and not some magical or judicious choice of variables. Weather and natural disturbance regimes are highly variable, and it is very difficult to separate effects of any anthropogenic disturbance (e.g., release of obscurant) from natural disturbance. Important examples include fire, flooding, drought, and pest outbreaks.

The selection of variables is directly relevant to study objectives and the nature of the ecological elements under investigation. If other factors are equal, selection of the variables with the least natural variation is highly desirable. Sampling logistics and difficulties should also be evaluated and determined if additional variability could be introduced as an artifact of either the sampling design or the sample collection method. Methods for measuring variables should be objective rather than subjective, because differences and measurement perception among observers is a serious source of bias.

Goldberg and Scheiner (1993) suggested that appropriate parameters to measure in ecological experiments, which can include analyses of effects of S/O materials, are:

- for individual-level responses to ecological or anthropomorphic stimuli: behavior, morphology, physiology, growth rate, age-dependent survivorship, and reproductive output or fitness
- for population-level responses: population numbers/biomass, and growth rates (e.g., relative or absolute density, biomass, cover, frequency, or other metrics)

- for community-level responses: taxonomic or functional group composition, dominance, and species richness.

Characteristics of populations and ecosystems that should be considered when designing time-dependent studies include changes in the following: (1) age distributions of species, (2) relative abundances of species, (3) migratory behavior, (4) stress rates, (5) spatial relationships among species, and (6) population gene pools (National Research Council [NRC] 1981). The capacity of an ecological system to store or detoxify chemicals should also be considered. Impacts of chemicals on ecosystems can be detected only if the natural structure and function of the system is well-understood (NRC 1981). Species-habitat relationships, patterns of change, and fluctuations or oscillations of populations are important parameters for impact assessment (Krzysik 1984, 1985).

Selection of Appropriate Experimental Units

Definition of experimental unit.

An experimental unit is the smallest subdivision of experimental material (or area) that can receive a given treatment. The number of experimental units used in a manipulative experiment is a major factor in determining the precision for estimates of variability among treatments. Sometimes in research with restricted budgets or resources, replicates within treatments are emphasized at the expense of using an adequate number of treatment comparisons. Although this strategy provides good estimates of within-unit variability, it compromises the ability to measure between-unit variability, which after all was the primary purpose of the experiment.

Representativeness.

The conditions being investigated in a study should be as similar as possible to the conditions under which the results will be applied (Cox 1958). The selection of experimental units that are representative of the species, material, or area to be evaluated is critically important to achieving results that can be applied in a real-world setting. Finding representative experimental units in S/O studies may be very challenging, because of the variety of conditions under which S/O are deployed, and because S/O are very sensitive to changes in external conditions, especially weather and terrain.

Spatial and temporal autocorrelation.

In many environmental studies, measurements are spatially or temporally related and correlated because experimental units that are close together in space and time may be more similar or related to each other than other, more distant units. Such a trend in variability is referred to as spatial autocorrelation, and measurements between adjacent areas have less variability than measurements between distant areas. If two measurements that are close in time have less variability than measurements that are farther apart in time, then temporal autocorrelation is present. Autocorrelations violate statistical assumptions, sometimes very seriously, because the experimental units are not independent of each other. Under these conditions, statistical inference may be tenuous or completely invalid.

Stratification.

If considerable variability exists across the range of a population, partitioning or grouping similar segments of the population in a sampling design can lower this variability. Such grouping of subpopulations by means of known characteristics is called stratification. In some experimental designs this may be an important way to increase statistical power and therefore sensitivity of an analysis. Examples of stratification occur in avian studies where the birds are classified as nestling, juvenile, and adult; or in regional studies where a geographic area is classified by elevation, topography, vegetative cover, classified ecosystems or plant communities, or some other parameter of interest. When sampling a species that occurs in many habitats over large landscapes, but nevertheless possesses a degree of habitat selectivity, important strata are habitat types or plant communities. Samples are randomly taken from each of the strata, often in a proportional manner, rather than being randomly selected from the population at large (e.g., if 16 percent of a region is mature upland deciduous forest, then 16 percent of the total samples will be taken from this area).

Replication.

The term "replication" is used in numerous contexts in statistical and ecological literature. In this report, replication refers to the assignment of more than a single sample to each treatment in the experimental design (Bender, Douglass, and Kramer 1989). For example, if two treatments are randomly assigned to eight experimental units, then Treatment 1 may be assigned to four experimental units and Treatment 2 may be assigned to four experimental units. This design has four-fold replication because there are four sets of the two treatments. If Treatment 1 were assigned to three experimental units, however, and Treat-

ment 2 assigned to five experimental units, then the design would have only three-fold replication because both treatments together, or repetition of the basic experiment, occurred only three times across all the experimental units. Replication can also be achieved by conducting all treatments together more than once (replication in time) or in more than one area (replication in space).

Replication is important because it increases the precision of an estimate by providing an estimation of experimental error, which is used to determine the significance of differences between treatment means (Hurlburt 1984; Bender, Douglass, and Kramer 1989; Underwood 1997). It is necessary to have a valid estimate of experimental error in order to conduct inferential analyses, because the error term is used for computing the correct probability for test statistics used in hypothesis testing. Replication of treatments is highly desirable because it provides a known probabilistic basis for determining true differences between treatments. A minimum of two replications is necessary to estimate experimental error. In practice, three or more replications allow the researcher to evaluate intrinsic differences between experimental units as well. In many environmental studies, however, replication may be uneconomic, very difficult, or impossible to achieve because of spatial scales or project magnitude, listed or rare populations, extreme logistical considerations, or replicates simply do not exist (e.g., a specific acidified lake).

Control sites.

To determine if S/O is affecting organisms, populations, or ecosystems, it is necessary to have an area where S/O have never been used to provide information on the natural variability of the experimental units. For example, an endangered plant population may naturally experience cyclical fluctuations in abundance and distribution. It would not be possible to separate the effects of S/O from natural cycles if the normal population patterns were not monitored during the same period the S/O study was being conducted.

Systematic Versus Random Sampling

Randomization is the assignment of treatments to experimental units in a manner that ensures that each experimental unit has an equal probability of receiving any given treatment. It also refers to the selection of sample sites or objects based strictly on random criteria. Randomness is a prerequisite for the estimation of experimental errors, which is the innate variability in experiments. For example, a field with varying levels of fertility may be divided into sections for an experiment. If randomization is used, each treatment in the study will have an equal chance of being assigned to a more fertile section or a less fertile sec-

tion. Randomization does not remove inherent properties of the experimental units, but it does introduce a "fairness factor" into the design by ensuring that no one treatment receives subjectively chosen favorable or unfavorable assignments (i.e., bias) (Bender, Douglass, and Kramer 1989).

If spatial patterns are present, systematic sampling can provide more accurate estimates of treatment differences than random sampling. Systematic sampling, however, cannot be used to provide an estimate of experimental error for hypothesis testing (Bender, Douglass, and Kramer 1989). A potential problem in a systematic procedure is that an undetected spatial pattern within the environment may coincide (or fail to coincide) with the spacing of the sampling points, resulting in sampling bias (Cox 1958), but this is highly unusual in practice. Systematic sampling is superior to random sampling and is required if the goal of the study is to assess spatial distribution patterns of organisms or chemicals in the environment. Systematic sampling is especially effective if populations exhibit aggregated or clumped patterns. Hairston, Hill, and Ritte (1981) found that systematic grid sampling correctly identified the spatial patterns associated with 17 of 22 species of soil arthropods; random sampling correctly identified the appropriate distributions for only 12 of 22 species.

Random sampling procedures are often desirable for environmental studies. Their major advantages include lack of bias, estimation of true experimental error, simplification of statistical assumptions concerning the population being sampled, and defensibility against criticism in legal situations (Borgman and Quimby 1988; Keith 1991). It is important to note that randomization assures a valid estimate of experimental error (Bender, Douglass, and Kramer 1989; Sokal and Rohlf 1995; Underwood 1997). Disadvantages of random sampling are cost, efficiency, and logistical considerations with respect to sample site location. Systematic sampling procedures require more careful preparation and justification than random procedures, especially if the systematic approach will be used to defend environmental decisions, but may offer substantial benefits in cost savings and interpretation (Borgman and Quimby 1988; Keith 1991).

Subjective or judgmental allocation of sampling units is tempting where costs, limited time, or other constraints are present. Although this approach may provide information about the effects of S/O on T&E species, these designs are biased and raise serious issues of validity and applicability of results. Data collected from sampling units that have been subjectively allocated are usually inadequate for resolving compliance/legal issues concerning S/O effects.

Nonhomogeneous Mixing and Deposition of Chemicals in the Environment

The release of military S/O into the atmosphere results in the formation of heterogeneous clouds with unpredictable dispersion characteristics. Farmer and Davis (1986) evaluated phosphorus mass concentration data acquired from several studies. They concluded that cloud homogeneity varied even over a distance of 1 m. Because the phosphorus clouds contained irregular areas of clear air ("holes"), low phosphorus concentrations, and high phosphorus concentrations ("hot spots") that continually changed in size and shape, they concluded that (1) no information was available to indicate what volume of air was appropriate to characterize the overall concentration of the cloud, (2) the chemical content of the clouds was highly time-dependent, (3) the distribution of the data became badly scattered as the clouds dispersed and holes became more numerous, and (4) sampling devices tended to undersample when concentrations were low.

These conclusions regarding the spatially unequal distribution of chemicals are consistent with other studies that evaluated dispersion and deposition characteristics of both military and nonmilitary chemicals. Haines (1993b) detected a 10-fold difference in fog oil deposition levels for two samplers placed side by side and exposed to ambient conditions for 24 hours. Harris (1984) found wide variations in 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) with concentrations ranging from 8.1 ppb to 57 ppb within a single square yard of soil. As these examples show, chemical concentrations in air and soil can vary widely even within a very small area; therefore, consideration of variability in samples should be a high priority when designing a monitoring strategy for S/O contaminants in the environment.

Independence of Experimental Units

Each of the experimental units should respond uniquely to the treatment being applied without being influenced by the response of the other units (Cox 1958). Satisfying this requirement ensures that the different treatment effects can be separated for evaluation. Independence of experimental units also eliminates crossover or lag effects for S/O impacts.

Sampling Units

Sampling units are the elements of the design that are actually measured. For vegetation sampling, a single sampling unit could be a leaf on a tree, the tree itself, or a collection of trees. The scale of the sampling unit depends on the na-

ture of the information desired, the spatial scale of the sampling design, and the type of design used to collect the sample.

Types of Designs

Completely Randomized Design (CRD).

A CRD uses a simple random selection procedure to select experimental or sampling units. In a CRD, each unit has an equal and known probability of being chosen for measurement. The units may be chosen either with or without replacement. If simple random sampling is performed with replacement, the units are returned to the group being sampled each time they are selected; therefore, the units have more than one chance of being selected. If simple random sampling without replacement is the method chosen, then each unit is withdrawn from the sampling pool as it is chosen; no unit can be selected more than once. Advantages of CRD are statistical validity, reliable estimates of experimental error, and straightforward analysis and interpretation of results (Krebs 1989). The major disadvantage is the lack of representative samples in spatial contexts or when appreciable heterogeneity is present. CRDs work best in situations where the experimental material is highly homogeneous, the effects of heterogeneity are not important to the objectives of the study, or the information needed to define strata is lacking.

Stratified Randomized Design (SRD).

With a stratified design, sets of treatments are randomly assigned within pre-selected strata (Krebs 1989). These strata could be habitat or ecosystem types, or groups resulting from a pilot study classification where experimental units within groups had lower variance than between groups for a selected criterion. A primary purpose of the stratified random design is to decrease experimental error resulting from natural environmental or organism variability by accounting for variance components extraneous to the study. In this way, analysis sensitivity is increased by increasing statistical power. The more quantitative information that is known about stratification, the easier it is to decrease Type II error (fail to reject null hypothesis when it is false). In experimental designs where Type I error (reject null hypothesis when it is true) is important, therefore, stratification provides the opportunity and statistical justification to *a priori* reduce α (P-value) in inference tests, while still maintaining low Type II error. In a military setting, stratification may be desirable when an S/O concentration gradient is present in the soil, or when the organisms being used as sampling units differ in age, sex, or size. Advantages of SRD are higher accuracy and lower variation across heterogeneous units. Additionally, the separate strata

can be analyzed as independent entities, which may sometimes increase the amount of information available from the data. Disadvantages are the extra effort required to identify strata and to allocate samples among strata, and more complex analysis and interpretation.

Systematic design.

A systematic design is based on using a sampling grid or other spatial sampling scheme in which sampling units are selected in sequential order at regular intervals (Krebs 1989). Typically, the location of the first sample is randomly selected, and all succeeding samples are taken at pre-determined, equally spaced intervals. Systematic designs can also be used in sequential studies where samples are taken at fixed time intervals, or where individual organisms are selected from a group according to a pre-determined sampling scheme (e.g., select every fifth mouse captured for analysis of tissue HC concentrations). Systematic designs are required for pattern analysis studies to determine the nature of spatial or temporal patterns.

Systematic-random design.

The systematic-random design is an excellent design for ecological field studies, because it fully utilizes the advantages and statistical properties of both systematic and random designs (Krzysik 1998a). The systematic component ensures sample representation and spatial coverage throughout the landscape of interest. This is particularly important when study sites are large and spatial heterogeneity is evident. The random component ensures sampling independence, objectivity, the avoidance of sampling bias, and correct estimates of experimental error. This is the design that was successfully used to assess the effects of landscape-scale military training activities on Mojave Desert vertebrate and plant communities (Krzysik 1984, 1985, 1994).

Factorial design.

Factorial designs are used in manipulative studies when the researcher desires to evaluate the interactions resulting from combinations of two or more treatments (Zar 1999). This design is motivated by, and is indeed mandatory for, assessing interaction effects among treatments. The factorial design may be incorporated within randomized or systematic designs. As an example, a 2×2 factorial design could be used to evaluate the effects of high and low fog-oil concentrations combined with high and low HC concentrations in a controlled experiment. The treatment combinations, or factorial arrangement of treatments, are shown in Table 1.

Table 1. Example of a 2 x 2 factorial design to evaluate the effects of high and low fog-oil concentrations combined with high and low HC concentrations.

	High HC	Low HC
High fog oil	High fog oil/High HC	High fog oil/Low HC
Low fog oil	Low fog oil/High HC	Low fog oil/Low HC

Repeated Measures Design (RMD).

In a repeated measures design, each experimental or sampling unit is sampled more than once. If the study is manipulative, then all units receive all treatments in random sequence. If the study is mensurative, then no treatments are applied, but each unit is measured for more than one trait or sampled more than one time. RMD may be used (1) when experimental manipulation is impossible, such as with human subjects, (2) when the amount of experimental material is limited, (3) when the researcher desires to use each experimental unit as its own control, (4) when there is a special need for the researcher to minimize between-unit variability, or (5) when the researcher wishes to measure S/O effects over time (Zar 1999). This design requires specific analysis procedures (Crowder and Hand 1990). An important advantage of this design is the minimization of within-treatment variability. The main disadvantage of RMD is that samples taken from the same units are autocorrelated and statistically dependent, and the design has limited interpretation and experimental flexibility.

Multi-stage design.

Multi-stage designs use a hierarchical grouping of units as successive stages in sample selection (Foreman 1991). Larger sampling units are selected in the initial sampling stage, and smaller sub-units are selected in successive stages. For example, the first stage of sample selection might entail randomly selecting one of several ecosystems to evaluate; the second stage would be random selection of individual plants within the ecosystem; the third stage would be the systematic selection of a certain number of branches on each plant; and the fourth stage would be the random selection of leaves on the selected branches.

Execution of Experiment: Sample Collection and Analysis

Pilot Study

A well-conducted pilot study is invaluable, often mandatory, for testing the feasibility of proposed field methods, discovering weaknesses in the sampling proto-

cols, assessing variability in data, determining sample sizes, identifying stratification schemes, assessing suitability of controls, collecting parameter values for the final experimental design, and developing satisfactory statistical analysis procedures. Elements of the design that should be critically evaluated during the pilot study include:

- locating sites with primary and secondary chemical constituents of interest
- developing criteria for sampling representativeness
- evaluating appropriateness of location and timing for sample collection
- determining equipment, materials, and methods needed for collecting field samples
- selecting methods for handling, transporting, and preserving biotic/abiotic materials and chemical samples
- using field determinations to assess degree of instability for chemicals
- identifying requirements for additional variables to be included in the design (Barcelona 1988).

The selection and characterization of control sites should also be conducted during the preliminary trials. The pilot study should involve 10 to 15 percent of the total sampling effort, while another 10 to 15 percent of the sampling effort should be reserved for resampling if necessary in the event of cross-contamination or other unanticipated problems (Keith 1991).

Quality Control

The quality of the data needed should be considered in the early stages of planning the sampling design. Data quality is based on the level of confidence required to meet study objectives. If the study is a preliminary exploration of contaminant extent and concentration, data quality criteria may be less stringent than if the study is being conducted in accordance with Federal, state, or other protocols to satisfy environmental regulations. Mandated studies must adhere to strict rules regarding sampling methods, transport of sample materials, and chemical laboratory procedures, or the data may be regarded as unacceptable (Keith 1991). The cost and effort involved with acquiring high-quality data may be beyond the amount budgeted for the effort. In such cases, potential compromises on data quality should be identified in advance, and either more funding allocated, objectives changed, the scale and/or resolution of the project adjusted within budget constraints, or the project should be dropped (Krzysik 1998a).

Considerable uncertainty exists in every part of the sampling process (Keith 1991). Efforts to control both experimental and procedural errors need to identify and address problematic areas in the field design layout, sample collection

techniques, sample transport process, laboratory analysis, and data reporting. Procedural errors such as sloppy or invalid field techniques, transposed or wrong numbers in recorded data, undetected cross-contamination, or deterioration of samples, are usually undetected and cannot be corrected. Such "hidden" mistakes may be common, are usually of greater magnitude than experimental errors, and can compromise the interpretation and validity of study results (Keith 1991).

Number of Samples

The feasibility of obtaining the sample size needed to achieve the desired level of precision should be evaluated in a pilot study or at least in focused field studies before initiating the project. This is important for sampling efficiency and economy, because field data collections and measurements are resource and time-consuming. Additionally, the amount of experimental material available may be limiting in S/O studies. Statistical analysis procedures and interpretations are simplified when balanced design is used with an equal number of samples measured for each treatment or characteristic of interest. When unequal sample sizes are unavoidable, robust analysis procedures should be selected to improve reliability. The adequacy of statistical power should be assessed before the project begins and reported in the results (Krzysik 1998a). The researcher also needs to consider in advance how to deal with missing data values resulting from samples that are lost, contaminated, or otherwise unusable.

Significance Level (α) and Statistical Power ($1 - \beta$)

Significance and power are related measures of the ability of a hypothesis test to predict the true condition of a population based on the data in the analysis. The relationship between these measures is shown in Table 2. The researcher should determine appropriate levels for α (alpha) and β (beta) in advance for an S/O study; once these values are known, the required sample size can be calculated.

Table 2. Relationships between the true condition of a population and the results of a statistical test.

		True condition of population	
Result of statistical test		H_0 true, H_A false	H_0 false, H_A true
	H_0 not rejected	Correct decision confidence level = $1 - \alpha$	Incorrect decision $\beta = P$ (Type II error)
	H_0 rejected	Incorrect decision $\alpha = P$ (Type I error)	Correct decision Power = $1 - \beta$

The confidence level for a statistical test describes the degree of certainty the researcher may place in the process used to generate the results of the test. The confidence level is expressed as the difference between 1.00 (perfect confidence) and α (the probability of committing a Type I error; see further explanation in the remainder of this section).

The significance level for the statistical test is α , which describes the probability of committing a Type I error, or the probability of rejecting a true null hypothesis (i.e., concluding from test results that a difference exists, when actually no difference is present). In statistical inference (hypothesis testing), the value of α must be set by the experimenter **PRIOR** to conducting the study, and it is referred to as an *a priori* rejection criteria.

β is the probability of committing a Type II error, or the probability of failing to reject a false null hypothesis (i.e., concluding from test results that no difference exists, when actually a difference is present). Statistical power ($1 - \beta$) is the probability of not making a Type II error. Report a power analysis with your data. Based on your sample size and the inherent variability in your data (error variance), how small a difference could you have detected as significant with the α value that you *a priori* selected (Krzysik 1998a). For an introductory discussion of statistical power, see Krzysik (1998a); for a comprehensive treatment of the subject, see Cohen (1988).

The probability of making a Type I error is inversely proportional to the probability of making a Type II error, so the consequences of making either should be considered carefully when designing a field study, especially if impacts on T&E species populations and habitats are being monitored. If a researcher makes a Type I error, he may conclude that T&E species or habitats are being affected by S/O when they actually are not, and the result may be undue restriction of military training activities. On the other hand, if a researcher makes a Type II error and concludes that S/O have no effect on T&E species populations and habitats when they actually do, a listed population could be impacted, and the installation would not be in compliance with the Endangered Species Act. It is the view of the authors that in conservation biology and in judgments and policies based on experimental results, it is desirable and prudent to make conservative decisions concerning species/population and habitat effects. Therefore, close attention and emphasis must be placed on not making Type II errors.

Minimizing Type II error is the same as increasing statistical power. Statistical power can be increased in four ways (Krzysik 1998a):

1. Use large or at least appropriate sample sizes, which increases degrees of freedom. Increasing sample size is the most important and usually the most feasible way of increasing power.
2. Design experiments with a small error variance (within population variance) and reduced confounding effects. This has the effect of producing a smaller denominator in the F-test; therefore, significance can be detected with smaller between treatment variance.
3. Increase the value of α . This is the usual alternative when sample size cannot be increased. Although this increases power and reduces the chances of making a Type II error, it increases the chances of making a Type I error. The trade-off is mutual when selecting between making Type I or Type II errors.
4. Increasing Δ (difference between population means that is *a priori* "considered" significant) increases power, because at any level of sampling variability, it is more reassuring to attribute significance to larger differences than to smaller differences.

P-Values

P-values or observed significance levels (OSLs) are the direct output of the analysis process in statistical inference. It is imperative that α is assigned **PRIOR** to the experiment or analysis. The P-value is used to determine whether to reject or not reject the null hypothesis. Once a statistical analysis is concluded, the P-value is compared to α . If the P-value is less than α , then the null hypothesis is rejected, and the alternative hypothesis is accepted with the potential for making a Type I error of α probability. When P is greater than α , then the experimental analysis has failed to reject the null hypothesis and, although the null hypothesis is "NOT PROVEN," it is accepted under the condition of the possibility of making a Type II error of β probability.

A common convention in biology has been to set α at 0.05 (a probability of committing a Type I error 1 out of 20 trials). There is no biological or statistical basis for the selection of $P = 0.05$, just "common usage" (Krzysik 1998a). Statisticians have argued (and published papers on the subject) for decades that the use of significance tests is over-emphasized in the reviewed scientific literature (see references in Krzysik 1998a). Nevertheless, biologists will find it convenient to use significance levels of 0.05, while scientists with less noisy (variable) data sets will use 0.01, and social scientists will use 0.1. In many research results, especially in mensurative studies, it may be more informative to simply provide analysis results in a P-value table, along with sample sizes and statistical power, without judgment of significance, so the reader is informed of the relative magnitude of the comparisons.

Physical Size (Volume or Mass) of Sample Unit

The physical size of the sample unit to be taken must be appropriate for the density and spatial distributions of the objects being measured (Barcelona 1988). For example, the mean and variance for the chemical concentration of a 50-g sample of soil may be quite different from that of a 500-g sample taken from the same site because of differences in spatially dependent processes such as degree of infiltration, bulk density, and obstructions such as rocks or roots. The optimal size and spacing of sampling quadrates depend on the size and spatial patterns of the objects that are being sampled (e.g., plant populations) (Kent and Coker 1992).

Length of Sampling Period

Consideration must also be given to the size of samples that exhibit time-dependent variation in the object being measured. A collection filter exposed to the air for sampling gaseous compounds will have increasing concentration with time. Such a filter may become oversaturated and fail to collect the full chemical load imposed upon it. In addition, chemical instability may result in degradation and reduced sample concentration over the course of the sampling period. Another factor to consider is the need to match the timing of sampling to reservoir turnover, release rates, or accumulation rates for each variable of interest (Green et al. 1991).

3 Statistical Analysis Considerations

Data Types and Data Quality

Data consists of numerical values assigned to some characteristics of the population of interest (Taylor 1990). The type of analysis procedure to be used may depend on the type of data collected. The amount of confidence that can be placed in the final results of an analysis depends a great deal on the quality of the data obtained.

Data Types

Discrete data are numbers with an exact value. Discrete data may consist of positive and negative integers, enumerators (counting numbers), or fractions that can be converted to finite decimal values. Examples of discrete data are the number of individuals in a population, number of paces from one location to another, and number of drops in a milliliter of liquid. The age of an organism is usually expressed as a discrete value. Fixed distances or times (e.g., points spaced every 0.5 meters along a transect; every 2 hours) are also considered to be discrete units.

Continuous data are numbers with measurement uncertainty associated with them. The uncertainty is caused by limitations in the ability of measuring devices to record values beyond a certain level of precision. Measurements of height, weight, and length are examples of continuous data. For example, height might be measured to the nearest meter, centimeter, or millimeter, depending on the resolution of the measuring device.

Interval data consist of numbers that are regularly spaced on an arbitrary scale with the location of zero defined by the researcher. Interval data can be discrete or continuous. Examples of interval data are time (seconds, minutes, hours, etc.), temperature (Kelvin, Celsius, Fahrenheit), compass degrees (North equals both 0° and 360°), and xy grid coordinates. A researcher may also create a specialized scale to describe characteristics of the data being collected (e.g., a scale of -10 to +10 to describe habitat desirability).

Percentile or ratio data are used to show relationships between two individual measurements, or between a single measurement and the sum of all measurements. Great care should be taken to ensure that a common reference base for comparison exists between the data values. Examples of valid percentile data are vegetation cover scores for quadrates or transects of fixed size (e.g., 0, 25, 50, 75, and 100 percent of the quadrate area or transect length is covered by vegetation), and percent slope data (e.g., a 5-ft vertical rise for every 100 ft of horizontal distance would equal a 5 percent slope). Percentile or ratio data with different reference bases should not be used for comparisons.

Qualitative data consist of non-numeric variables that convey information about the object under study. Examples of qualitative data are gender (male, female), colors of the rainbow (red, orange, yellow, green, blue, indigo, violet), and health status (healthy, nonhealthy). Qualitative data are often recoded to numeric values for analytical purposes.

Rank or ordinal data consist of numeric or non-numeric values arranged in a definite order. Rank data can be in ascending order (smaller to larger values) or descending order (larger to smaller values). Examples of rank data are relative size or amount (small, medium, large), habitat quality (poor, fair, good, excellent), vegetation density scores (5 = dense vegetation, ..., 1 = sparse vegetation, 0 = no vegetation), and species association scores (-1 = species are never found together, 0 = species are neutral with respect to co-occurrence, 1 = species are always found together). In addition, discrete or continuous data can be ordered and assigned a ranking for certain kinds of analyses.

Categorical data. Sets of discrete or continuous data may be grouped and analyzed by categories. Examples of categorical data would be elevation above sea level (Category 1 = 0-100 ft, Category 2 = 101-200 ft, etc.), avian life stages (nestling = 0-50 days, fledgling = 51-90 days, juvenile = 91-365 days, adult = 365+ days), or pH levels (very acidic = pH 1 to pH 3, moderately acidic = pH 4 to pH 6, ..., very basic = pH 11 to pH 14).

Binary data are a special case of categorical data consisting of 0 and 1 values assigned to distinguish between two mutually exclusive conditions. Binary data are most often used to indicate presence (value = 1) or absence (value = 0), or to indicate if a particular condition is true (value = 1) or false (value = 0). Contrary to popular belief, ordinal, categorical, and binary data may be superior to continuous metric data in many ecological contexts, especially in multivariate analysis (Krzysik 1987).

Data Quality

Ensuring high data quality is critically important to the success of a research project. Taylor (1990) stated,

It is almost useless to apply statistical techniques to poorly planned data. This is especially true when small sets of data are involved. In fact, the smaller the data set, the better must be the preplanning activity. Any gaps in a data base resulting from omissions or data rejection can weaken the conclusions and even make decisions impossible in some cases.

Factors that affect data quality and its subsequent analysis include: reliability, representativeness, inherent variability, bias, procedural errors, precision, accuracy, sensitivity, and outliers.

Reliability is data quality that can be documented, evaluated, and believed (Taylor 1990). The experimental or sampling design should be completely and carefully documented so that the steps used to collect data are clearly outlined. The assumptions used in developing the data collection protocols, the sampling procedures used, quality control procedure implemented, and any problems encountered during the sampling process should be included in the documentation. Peer review of the design before it is implemented in the field is highly recommended. The peer review should include the input of one or more professional statisticians and expertise in the field of investigation — especially when field studies are involved.

Representativeness is simply meeting the condition that sampled sites or objects are representative of the population of interest. This data quality is discussed in Chapter 2 in the section "Determination of True Population To Be Sampled."

Variability or random errors are the difference between the true value of a parameter and the values of each measurement used to estimate the true value. Inherently, some environmental variables are more variable (noisy) than others. Random errors associated with taking many measurements will have an average of zero in the long run (Taylor 1990). For example, if several measurements using the same scale are used to determine the mass of a fish that is exactly 2.00 kg, the actual measurements recorded might be 1.94, 2.01, 2.17, 2.00, 1.87, and 1.96 kg. The differences between each measurement taken and the true mass of the fish are the random errors. Environmental data contain numerous sources

of variability. Sources of such variability must be adequately identified and quantified for sampling efforts to be successful. Triegel (1988) noted:

In the initial stages of planning a sample collection program, identification of the potential sources of variability is critical. The nature of the variability may affect the number of samples to be collected, the method(s) of collection and analysis, and the overall design of the sampling program. The identification of the sources of variability and bias before starting field operations may eliminate the use of inappropriate collection and analytical methods or sampling intervals.

Bias is defined as systematic error associated with a given measurement process which always has the same sign and magnitude (Taylor 1990). An important source of bias is personal researcher or surveyor subjectivity in collecting data, making measurements, or selecting sites or individuals/objects. Other biases include: unrepresentative sampling, degradation of chemical compounds between the time of sampling and laboratory analysis, improperly calibrated instruments, and protocol mistakes. For example, a mass balance that measures 2 g short of the true mass of a sample will give consistently lower values for all samples measured. An instrument that has been calibrated at one temperature but used at a different temperature may introduce bias into the results. If time were the measurement of interest, then a clock which runs 10 minutes ahead of the true time would have a positive bias; a clock which runs 10 minutes behind the true time would have a negative bias. See Green (1979) for discussions of bias.

Procedural errors are errors that are the result of poorly executed experiments, unstable measurement systems, or poor execution of data measurement or collection. These errors are not statistically manageable and can invalidate an otherwise good research design or sampling method (Taylor 1990; Lessler and Kalsbeek 1992).

Precision, usually expressed in terms of standard deviation, has been defined as a measure of mutual agreement among individual measurements of the same property (Smith et al. 1988). The final precision of estimated treatment effects depends on several factors, as follows (Cox 1958): (1) the intrinsic variability of the experimental material, (2) the accuracy of the sampling effort, (3) the number of experimental units measured, (4) the number of subsamples taken from each experimental unit, (5) the nature of the experimental design and sampling methods, and (6) the method of statistical analysis.

Accuracy describes the magnitude of systematic error present in a series of measurements (Keith 1991). If the systematic errors associated with the measurements are small, the measurements have high accuracy. For instance, if the true temperature is 21.000 °C, and a thermometer reading is 21.005 °C, the thermometer has high accuracy. If the thermometer reading is 28.000 °C, the thermometer has low accuracy. Such instruments may state their accuracy as a percentage of their range, e.g., -50 to +200 °C ± 1% (i.e., 1% of the 250 °C range, or 2.5 °C).

Sensitivity is the ability of an experimental design to detect true differences if they exist (Bender, Douglass, and Kramer 1989), and is directly related to statistical power (Cohen 1988). It is defined as the inverse of the standard deviation for the difference between two means. In other words, if two or more experimental designs could be used to estimate the means of a variable of interest under two different sets of conditions (the difference between a smokes area and a control area, for example), the design that can detect the smaller difference between the two means is the more sensitive design.

Outliers are observations that deviate substantially from the majority of the observations in a data set. They can have a considerable effect on the results of an analysis procedure and could potentially cause a researcher to draw erroneous conclusions from the data. If outliers are detected in a data set, the researcher should consider how the presence of the outlier will affect analysis results. Conducting the analyses with and without outliers and evaluating the difference that outliers make is highly recommended. Outliers should not be arbitrarily excluded from an analysis; rather, an assessment of their influence should be undertaken, and the decision to include or exclude them should be based on the extent of their effect on the results. The reason that the observation is an outlier should also be considered — is the outlier a result of natural variability in the data, observer error, an abnormality in the conditions present at the time of the measurement, or some other factor? Such an evaluation of unusual data may provide valuable insight into the data set as a whole.

Approaches to Statistical Analysis

Statistical analysis consists of at least six general approaches: estimation, descriptive statistics, exploratory data analysis (EDA), inference, modeling, and spatial analysis (Krzysik 1998a).

Estimation

The most common example is estimating the mean and associated precision in a population of interest. The precision in estimating the mean (or another statistic) depends on inherent variability in the population and the sample size used to estimate the statistic under investigation. Statistical precision is called error and is expressed as standard deviation, standard error, confidence interval, or coefficient of variation. Estimation directly leads to descriptive statistics.

Descriptive Statistics

Descriptive statistics are generally summary statistics for all the primary parameters or variables in the project, generally stratified by spatial, temporal, or user-defined classes. Summary statistics are provided by all statistical analysis packages. Graphical outputs and displays are indispensable components of descriptive statistics. The important foundation in the philosophy and techniques of data display has been the work of Tufte (1983, 1990). Practical guidance for using graphics effectively can be found in Chambers et al. (1983) and Cleveland (1993).

Measures of central tendency or location.

Measures of central tendency or location provide estimates of the central or middle value for a set of measurements. Different measures of central tendency are used, depending on the distribution of the data, the presence or absence of outliers, and degree of symmetry. The most common measures of central tendency are the arithmetic (sample) mean, geometric mean, median, and mode.

Arithmetic mean. The sample mean, also called the average, is a measure of the central value for a set of measurements. It is calculated as the sum of all measurements divided by the number of observations. The mean is an effective measure of central tendency only if the underlying distribution of the data is symmetrical. The mean is very sensitive to outliers, so even a few unusual observations may unduly influence the results.

Geometric mean. The geometric mean may be used when (1) the parameter of interest is a rate or ratio, or (2) when a measurement taken in one time period is dependent on a measurement taken in a previous time period. For example, if a researcher wishes to evaluate the average population growth rate of purple balduina (*Balduina atropurpurea*) over a 5-year period, then the geometric mean of population growth rate would be a more appropriate statistic than the arithmetic mean.

Median. The median is the middle value in an ordered set of numbers if the number of observations is odd. It is the average of the two middle values of ordered numbers if the number of observations in the set is even. Half of a sample has values larger than the median and half has smaller values. The median is a more stable measure of the central value for a series of measurements if outliers are present or if the distribution is skewed (asymmetrical). A common example of this is that the median is a more meaningful metric than the mean for characterizing the price of housing in any locality. This is because the cost distribution has a highly skewed tail for expensive homes (i.e., the highest priced homes can be significantly above the mean, while the cheapest homes can only approach "0" to some rational finite cost). This distribution inflates the value of the mean relative to the median.

Mode. The mode is the most frequently occurring value in a set of numbers. In the set {14, 25, 18, 17, 14, 65, 11}, for example, 14 is the mode because it occurs more often than the other numbers.

Measures of dispersion.

Measures of dispersion provide information about how far the measurements extend away from a central value (i.e., variability or scatter). Given three sets of numbers $A=\{60, 60, 60, 60, 60\}$, $B=\{20, 40, 60, 80, 100\}$, and $C=\{58, 59, 60, 61, 62\}$, one can see that the mean for all three sets is 60, but the extent to which the numbers in each set differ from 60 is quite different for the three sets. The measures of dispersion most commonly used to evaluate this deviation from the mean are the range, variance, standard deviation, standard error, coefficient of variation, and confidence interval for the mean.

Range. The range is the largest value of a set of numbers minus the smallest value. It is a measure of the extent of variation in the data. For a set of numbers {2, 33, 14, 28, 43} the range would be $43 - 2 = 41$. The range is the simplest measure of dispersion to calculate, but contains limited information about the nature of the scatter.

Variance. Variance is a weighted measure of distance between the observations in a sample and the sample mean. Since variance is a squared value, it is always positive. The larger the variance, the greater the overall distance between the measurements and the mean for the sample.

Standard Deviation (SD). The SD is the square root of the variance. It has the advantage of being expressed in the same units as the original measure-

ments. The SD is usually the most effective way of showing variability in a given data set.

Standard Error (SE). The sample SE is the SD divided by the square root of the number of observations. SE closely reflects the precision in estimating the mean. It is used to indicate the relative precision of the SD when the measurements come from several sets of observations rather than from individual observations. For example, an SD calculated from a sample of 5,000 observations would be a better estimate of the true dispersion of data about the mean than an SD calculated from a sample of 10 observations. This finer scale of precision is reflected by the SE. The value of SE compared with SD is directly related to sample size. As sample size increases, the SE decreases in relation to SD.

Coefficient of Variation (CV). The CV is a relative measure of the spread of the data. To use it effectively, the researcher should be familiar with related data to determine if the spread is unusually large or small compared with the other data sets. The coefficient of variation is defined as the SD divided by the mean.

Confidence Interval (CI) for μ . Sometimes a researcher may wish to express the variability of the measurements about a mean as a range of numbers rather than as a single number. One way to accomplish this is to use the CI for the mean. The result is expressed as (LL, UL), where LL is a number that indicates the lower limit of confidence for the data and UL is the upper limit. The value of the CI is: $CI = \text{mean} \pm (SE \times t_{\alpha})$, where t_{α} is the value from a t-table at the α level. For example, when $\alpha = 0.05$, $t_{\alpha} = 1.96$. This means that, for normally distributed data, 95 percent of the data lies between $-1.96SE$ and $+1.96SE$ of the true mean. See Sokal and Rohlf (1995) or any basic statistics textbook for more information.

Exploratory Data Analysis (EDA)

EDA is an important class of statistical analysis that has not been fully appreciated despite an excellent and technical foundation by Tukey (1977). EDA has also been called Initial Data Analysis (IDA) by Chatfield (1988), who concludes that the process is indispensable and required by the statistician to get a feeling for the data. The routine use of EDA has become a current reality because of the power of modern microcomputers and the availability of interactive graphics and extensive graphics output options in microcomputer statistical software packages (e.g., SPSS, SYSTAT, S-PLUS, MINITAB, SAS).

Interactive graphics enables rapid examination of data patterns and trends from scatterplots of raw data, transformed or rescaled data, or residuals (Chambers et al. 1983; Cleveland 1993). The scatterplot matrix is an important procedure. For example, if there are 10 variables in the data set, the scatterplot matrix routine produces a single plot containing 100 subplots of each combination of the 10 variable pairs plotted against one another. The plots above the diagonal are the same as the plots below the diagonal except that the ordinates and abscissas of all paired variables are interchanged.

Graphs.

Ellison (1993) gives a good overview of several types of graphical displays for data analysis, and of the strengths and weaknesses associated with such displays. Graphical displays commonly used to investigate patterns in data include bar charts, pie charts, and scatterplots. Other graphics that can be used to display summaries of statistical information for pattern analysis and comparisons are frequency histograms, box-and-whisker plots, and stem-and-leaf plots. Probability plots provide visual estimates of whether or not data fit a given distribution (e.g., normal probability plots are used to evaluate whether data are distributed according to a Gaussian distribution). Additional suggestions for presenting data are demonstrated by Green (1979).

Statistical distributions.

A statistical distribution, or probability distribution, is an arrangement or pattern of data values around a central value which can be described by mathematical functions, called probability density functions. Generally, the minimum amount of information needed to characterize a distribution will include the mean, sample standard deviation, and the number of samples used in the calculations (Taylor 1990). Some common distributions in ecology are the binomial, negative binomial, Poisson, normal, chi-square, exponential, and lognormal. Refer to Beyer (1988) or Hastings and Peacock (1975) for descriptions and properties of these distributions.

Many hypothesis tests are based on the assumption that the data follow a normal (Gaussian) distribution. Such tests fall into the category of parametric analysis techniques. Since many kinds of ecological data violate this assumption, the appropriateness of using such data for inferential statistics should be determined prior to analysis. Some types of data can be transformed mathematically to approximate a normal distribution; however, problems with interpreting the transformed results may arise. Nonparametric tests are statistical tests that make no assumptions about the distribution of the data. Nonparamet-

ric tests should not be used, however, as an excuse for poorly designed or executed research studies or with ill-behaved data sets. Researchers all too often rely on these tests as a last resort to justify the use of poor data (Krzysik 1998a). Additionally, researchers may not be aware that these tests are subject to the same limitations of asymptotic behavior, reasonable sample sizes, and sample independence as are parametric tests (Krzysik 1998a). It is probable that a majority of the research community believes that nonparametric methods possess low statistical power in contrast to parametric tests, but in reality the difference is not practically significant (Krzysik 1998a).

Inference

Inference or hypothesis testing is probably the most familiar use of statistical analyses. Inference is applied by the investigator to decide if the observed difference in a test statistic (e.g., mean) between two or more populations should be considered different or due to chance at some *a priori* set probability. The question is posed as a null hypothesis to falsify (null hypothesis: populations are homogeneous). If no difference exists between two or more populations, what is the probability of selecting samples with differences as large as or larger than those observed? This probability is the familiar P-value or α . If probability is very small, then one concludes that the differences are unlikely to be due to chance, and there is a statistically **significant** difference in the populations (null hypothesis rejected) at the P-level. If probability is large (observed differences may be due to chance alone), then either the populations are homogenous at the P-level, or the statistical power of the test was too low (i.e., some combination of small sample size, high natural variability, or the "difference" selected to assess significance was too small). It is imperative to remember that the null hypothesis can never be proved correct, but can only be rejected with a known risk of being wrong.

Modeling

Modeling represents the efforts to verify that experimentally derived data fit specific mathematical models related to biological, physical, geological, or chemical phenomena or processes. The most common example in statistics is linear regression. Do the data fit a straight line? Of course, any kind of polynomial curves in any dimensions can be equivalently modeled, but with much more difficulty. The four main strategies in model building are model formulation, model estimation or fitting, sensitivity analysis, and model validation. Model validation includes the familiar:

$$\text{Experimental data} = \text{mathematical model} + \text{residuals.}$$

For further analysis the residuals can be subjected to standardization (homogeneous variances), their distribution can be examined by using probability plots, plotting residuals versus selected variables, or the residuals can be subjected to further analysis or modeling. The analysis of residuals may provide valuable insight into a very important facet or unexpected behavior of the model.

Spatial Analysis

Spatial analysis has developed quite independently from mainstream statistics and even has its own terminology. Spatial statistics is based on data that are georeferenced. In other words, data points are referenced to two or three dimensional occurrences in space. Spatial statistics and its toolbox, therefore, are closely associated with geographic information systems (GIS), and the analysis, description, projection, or display of point, vector (line segments), and polygon patterns of landscape elements. For an introduction to GIS and its literature see Krzysik (1998b). An important capability of spatial statistics is the interpolation and smoothing of spatially explicit field-collected data for prediction, visual interpretation, and demonstration (Krzysik 1998b). As a direct result of this capability, an important application is the use of Thin-Plate Splines for modeling and monitoring the distribution and density patterns of T&E populations (e.g., the desert tortoise; Krzysik 1997). Spatial statistics is computer intensive and was once the domain of mainframe and minicomputer workstations, but is rapidly gaining popularity because of the widespread availability of "inexpensive" high-powered microcomputers. See Krzysik (1998a) for fundamental references.

Univariate Statistics

Univariate procedures are statistical analyses that contain only a single dependent or response variable, and one or more independent variables or predictor variables (simple linear regression). Additionally, some univariate statistics may have two independent variables (e.g., bivariate correlation). Both parametric and nonparametric methods are discussed.

Parametric Methods

Parametric analysis procedures are based on three primary and important assumptions, listed here in order of their importance (Krzysik 1998a):

1. Observations are independent of one another: random observations, sampling or experimental errors are independent, and avoidance of sampling or experimental bias
2. Populations (comparisons) possess homogeneous variances (residuals or data scatter)
3. The data from population samples or observations are normally distributed.

These assumptions can formally be tested, but typically they are not. Goodness-of-fit and normality tests and calculations of skewness and kurtosis (see glossary) are generally available in all basic statistical packages. Bartlett's test assesses homoscedasticity (occurrence of equal variances among treatment groups), but its practical value has been questioned (Harris 1975), and it is unduly sensitive to non-normality. Cochran's test (1951) uses the ratio of the largest variance to the sum of all sampled variances as a test statistic, and may be the most desirable test for the presence of excessive heteroscedasticity (Underwood 1997). Sampling independence is usually difficult to detect, and is directly related to a proper experimental design. In some cases, correlational tests or the examination of scatterplots of the raw data may detect it. Parametric methods are generally considered robust with respect to these assumptions, especially assumption number 3, when sample sizes are reasonable (e.g., 20 to 30) and because of the central limit theorem, particularly when the raw data have been properly transformed (Krzysik 1998a). However, assumption number 1 can often lead to invalid statistical inference, even with large sample sizes. Transformations only apply to assumptions 2 and 3.

Parametric methods are the well-known statistics taught in introductory statistics courses, and represent the methods most frequently used in biological research. Familiar examples include: analysis of variance (ANOVA), analysis of covariance (ANCOVA), correlation analysis, and regression models. Linear regression belongs to the family of generalized linear models (GLM), and ANOVA and ANCOVA are special cases of linear regression. Nonlinear or polynomial regression and multiple regression (more than one independent or predictor variables) are extensions of the basic model. Fundamentals of GLM and modeling are provided in McCullagh and Nelder (1983), Cullen (1985), Neter, Wasserman, and Kutner (1985), and Dobson (1990).

Milliken and Johnson (1984, 1989) present practical approaches and methods of data analysis for experimental designs and parametric data that are plagued with the well-known problems associated with field data: failures in assumptions, unbalanced designs, lack of replication, repeated measures, multiple comparisons, outliers, and missing data.

Analysis of Variance (ANOVA).

ANOVA is a statistical analysis procedure that examines and explores sources of variation in sample data. Details in the theory and application of ANOVA can be found in any basic statistics textbook. Particularly useful are Myers and Well (1995), Sokal and Rohlf (1995), Underwood (1997), and Zar (1999). ANOVA tests the null hypothesis that there is no difference in a variable of interest among two or more populations classified by one or more criteria. These criteria are called factors and are commonly referred to as treatments and controls. Valid replication and interspersion for all treatments are critical (see Krzysik 1998a). Essentially, ANOVA uses the F-test statistic to assess if the magnitude of the ratio of the variability between treatments to variability within treatments is so high that it is unlikely to occur by chance alone at an *a priori* selected α error rate, and the null hypothesis is rejected. Conversely if the ratio is small, the observed ratio of variances could have occurred by chance, and the null hypothesis cannot be rejected.

In the simplest case of one-way ANOVA, the variable of interest is tested in two or more populations (groups) that are classified by a single factor or treatment. When there are only two groups, the analysis is called a Student's t-test. The Student's t-test for comparing a single mean to a known population value is used if the researcher wishes to compare sample data from a population to a standard reference value. The Student's t-test is appropriate if (1) only one treatment level is used, (2) one response variable is measured, (3) the data represent random sample of size n , and (4) the sample data come from a population with a normal distribution (Steel and Torrie 1980).

When more than one factor is present, the analysis is called a factorial ANOVA. A factorial ANOVA design is much more powerful than using separate one-way ANOVAs (i.e., one for each factor). In the case of a two-factorial design, for example, not only can the main effects of factor A and factor B be assessed, but their interaction effects ($a \times b$) can be as well. Sokal and Rohlf's (1995) classic three-factor experimental example measured the survivorship of minnows (variable of interest) at five different cyanide concentrations (factor A), at three different temperatures (factor B), and at three oxygen concentrations (factor C). Thus ANOVA with just a single variable can be extended to test many factors, but the inherent complexity of ever increasing multiple interaction effects make interpretation tenuous. For example, with only three factors the possible interaction effects are: $a \times b$, $a \times c$, $b \times c$, and $a \times b \times c$.

Nested ANOVA experimental designs are particularly important for ecological field studies because they help to achieve valid replication and interspersion of

study plots or samples (Krzysik 1994, 1998a). Nested ANOVAs can be applied to any number of factors (including one) and refers to the provision of two or more randomized subgroups within each population or primary group.

Potvin (1993) distinguished between fixed and random factor effects for ANOVA. (See also Krzysik 1994, 1998a.) Fixed factors are drawn from samples that represent specific levels of interest deliberately chosen by the researcher, while random factors are drawn from samples that represent all conceivable levels for the entire population. For example, in a fixed factor effects design, a researcher may choose to investigate the effects of white phosphorus concentration levels of 2.5 and 3.5 ppm on triglyceride metabolism in wood storks. In a random factor effects design, however, the researcher would choose to study the potential for metabolic changes in wood storks at concentration levels of white phosphorus randomly selected from all possible levels. If a treatment level effect is fixed, then conclusions cannot be generalized beyond the levels used in the study (Potvin 1993).

Balanced ANOVAs are required to obtain unambiguous interpretation of interaction effects and overall significance. Balanced means that there are equal observations in each experimental treatment. Balanced designs cannot always be used for the practical collection of ecological field data. Shaw and Mitchell-Olds (1993) review ANOVA for unbalanced designs and provide guidelines for the analysis of fixed effects models.

In repeated-measures analysis, the same experimental or sampling unit is measured for more than one variable, or the same unit is measured more than one time. Repeated-measures analyses are recommended for evaluating trends over time, for assessing pre-impact and post-impact effects of S/O in acute and chronic bioassay studies, and for monitoring very small populations. Repeated-measures analysis represents an important statistical protocol that can be analyzed as a univariate, randomized complete-block or split-plot ANOVA design, or as a multivariate ANOVA (MANOVA). MANOVA is used to simultaneously assess the relationships between one or more treatments (independent variables) and two or more dependent variables. Crowder and Hand (1990) and Stevens (1996) have more details. Univariate designs are explained in basic texts such as Sokal and Rohlf (1995) or Zar (1999). A repeated-measures design, with the use of a unit as its own control, improves statistical power, sometimes dramatically, because variability among subjects due to individual differences is removed from the error term in variance comparisons (Stevens 1996). Smaller error variance terms (denominator in the F-test ratio) can detect significant differences at a given value of alpha with smaller between-variance components. Additionally,

fewer experimental subjects are required than in completely randomized designs.

A great deal of controversy exists over the relative merits, preference, and selection of univariate or multivariate repeated-measures approaches (von Ende 1993; Stevens 1996).^{*} Barcikowski and Robey (1984) and Stevens (1996) suggest that both univariate and multivariate analysis be conducted to determine if the two approaches differ in detecting treatment effects. Stevens (1996) further recommends adjusting the degrees of freedom by averaging ϵ ("error" or "residual") from both the Greenhouse-Geisser and Huynh-Feldt corrected probabilities, and using an $\alpha = 0.025$ for both univariate and multivariate tests.

For exploratory analysis, both univariate and multivariate models should be analyzed and compared, and dependent variables should be analyzed both separately and in combination. As an additional suggestion, raw data should be transformed to stabilize variances and distributions (Krzysik 1998a).

Analysis of Covariance (ANCOVA).

ANCOVA is a statistical procedure that encompasses both ANOVA and linear regression. ANCOVA is used when the means of two or more populations are being compared, but the variable of interest is confounded by another variable that may or may not have the same effect on the populations. This variable is called a covariate, and linear regression is used to "adjust" for its influence. One of the

Mead (1988: Section 14.5) and Underwood (1997: Section 12.5) discuss the problems and assumptions with use of time as a within-subject factor. They favor multivariate approaches, because their main concern is nonindependence of temporal measurements. Additionally, an important consideration is that MANOVA requires fewer assumptions of homogeneity of variances and covariances across subject trials and factors (Wilkinson, Blank, and Gruber 1996). For example, the important assumption of sphericity, which requires the variances of the differences of all pairs of repeated-measures being equal, is not necessary (Stevens 1996). However, MANOVA also requires adequate sample sizes; for repeated-measures, sample size must be higher than $k + 10$, where k is the number of levels in the within-subjects measure (Maxwell and Delaney 1990). General consensus has been that univariate approaches, while having higher power, require more rigorous assumptions (Gurevitch and Chester 1986). However, ANOVA and MANOVA are robust to deviations from normality, and heterogeneous variance — covariance structure is more important (Underwood 1997). Box's M test (Box 1949) can be used to test if the covariance matrices of dependent variables are homogeneous across all level combinations of between-subjects factors. Box's test is very sensitive to non-normality (Stevens 1996), inspiring further confidence in normality assumptions. Shapiro-Wilk's test (Shapiro, Wilk, and Chen 1968) should be used for formally testing the assumption of normality. Levene's test (Levene 1960) should also be used to test for equality of error variances of each dependent variable among groups.

important tests of ANCOVA is to assess if the regression lines of the populations under analysis possess similar slopes.

The most important uses of ANCOVA are (Steel and Torrie 1980): (1) control error and increase precision, (2) adjust treatment means of the dependent variable for differences in sets of values of corresponding independent variables, (3) assist in data interpretation of treatment effects, (4) partition total covariance into components, and (5) estimate missing data. Snedecor and Cochran (1989), Sokal and Rohlf (1995), and Underwood (1997) present good treatments of ANCOVA.

The concept of ANCOVA can most readily be shown by an example. Suppose we want to test the effect of altitude on egg production by a given species of salamander. The hypothesis could be that, because lower elevation populations are exposed to milder temperatures (and therefore longer seasonal activity and invertebrate prey availability), lower elevation populations (for a given body size) should produce larger egg clutches. This is a straightforward ANOVA problem. It is also known, however, that body size directly affects egg production (larger salamanders have larger egg clutches), and altitude may also affect population body size. Therefore, clutch size is potentially determined by two factors: elevation and a linear relationship (after transformation) with body size. ANCOVA is the appropriate statistical model to use in this case, where body size is treated as a *covariate*, essentially "correcting" for this factor when the interest is the variability in egg production between two elevations.

In certain types of manipulative studies, direct methods for increasing precision and removing bias through the experimental design are not possible. In such cases, ANCOVA may allow the researcher to control variability due to experimental error by using statistical analyses procedures after the data are collected (Winer 1962). The assumptions for ANCOVA are the same as for analysis of variance, plus the three assumptions listed below (Stevens 1992). Violations of any of the following three assumptions will seriously affect the validity of test results:

1. A linear relationship exists between the dependent variable and the covariates (data transformations, such as the logarithmic, can change a nonlinear relationship into a linear one)
2. The slope of the regression line is the same in each group — tested statistically based on the data
3. The covariates are measured without error.

Regression and correlation analysis.

Regression analysis. Regression analysis represents the important model:

$$y = m_i^j x_i^j + b_i^j + \text{error}$$

When I = 1, j = 1: Simple linear regression.

When I = 1, j = 1, 2, ..., k (k usually < 4): Simple polynomial or nonlinear regression.

When I = 1, 2, ..., n, j = 1: Multiple linear regression.

When I = 1, 2, ..., n, j = 1, 2, ..., k (k usually < 4): Multiple polynomial regression.

Although regression analysis is well covered in the fundamental texts referenced earlier in this section, other valuable texts include: Draper and Smith (1981); Montgomery and Peck (1982); Cohen and Cohen (1983); Neter, Wasserman, and Kutner (1985); and Chatterjee and Price (1991).

Other regression analyses that have extensive applications in ecology are logistic regression and locally weighed scatterplot smoothing (LOWESS) regression (Trexler and Travis 1993). Logistic regression deals with dichotomous (bivariate) or polychotomous dependent variables and transforms the data to model binomial or multinomial distributions. LOWESS models the relationship between a dependent (response) variable and independent variables under the assumption that neighborhood values of independent variables within a given range are better indicators of the dependent variable in that same range.

Correlation analysis. The correlation coefficient, ρ , is a measure of the strength of the relationship between two variables. The value for ρ ranges from -1 to +1. If $\rho = 0$, then the variables are not correlated. If $\rho = +1$, then the variables are perfectly and positively correlated; as the value of one variable increases, the value of the other variable increases. If $\rho = -1$, then the variables are perfectly and negatively correlated; as the value of one variable increases, the value of the other variable decreases.

The coefficient of determination, r^2 , is the square of the correlation coefficient. It describes the amount of variation in the dependent variable that can be attributed to the independent variable.

Nonparametric Methods

Nonparametric statistics (NPS) are also called distribution-free statistics, because they make no assumptions about test statistic distribution (e.g., a normal distribution) and other behaviors. They are also required for the analysis of ordinal or categorical data. Many researchers believe that nonparametric methods possess low power in contrast to parametric tests. In reality, the difference is not practically significant (Siegel 1956; Hollander and Wolfe 1973; Noether 1987). What is not always appreciated, however, is that NPS, like parametric tests, are also subject to the same two most important limitations and violations of statistical analyses — nonindependence of sampling errors (the need for random sampling) and the loss of statistical power when sample sizes are too small (e.g., Box, Hunter, and Hunter 1978; Stewart-Oaten 1995). Additionally, high heterogeneity among sample variances also can affect these tests. The chi-square test is the best known nonparametric test and perhaps the most misused. Siegel (1956), Hollander and Wolfe (1973), and Conover (1980) are fundamental texts for nonparametric analysis. Siegel's book presents a very useful classification table of nonparametric methods to guide the user. The basis for the classification is number of sample comparisons and data scale (nominal, ordinal, or interval).

Academic controversy in the literature concerns the use of NPS statistics. The basic argument goes like this:

Proponents — Because NPS possess almost the same power as parametric tests and avoid the assumption that the data are normally distributed, while environmental data are usually non-normal, NPS should be more routinely used in ecological research and monitoring (e.g., Potvin and Roff 1993).

Opponents — Parametric tests are more powerful and reasonably robust to the stated assumptions. NPS do not help with the serious violations of independence and heteroscedasticity. Both approaches are sensitive to small sample sizes and strongly unbalanced data. The assumption of normality is the least stringent assumption and effectively treated with appropriate transformations. NPS have their own assumptions, which are not often appreciated. NPS should not be used (as it is sometimes) as an alternative to poorly conceived experimental or sampling designs or poor field procedures or just poor data (e.g., Johnson 1995; Smith 1995; Stewart-Oaten 1995; Underwood 1997).

A survey of some common nonparametric tests follows.

Chi-square test. The chi-square test is the most familiar and frequently used nonparametric test. Biology students receive an early exposure to it in introductory genetics courses. The value of the chi-square test is its potential broad applicability, including its compatibility with nominal scale data. This test is used to determine the significance of the differences among N independent groups when the research data consists of frequencies in discrete categories (either nominal or ordinal).

Cochran's Q-test. Cochran's Q is a method for testing if three or more matched sets of frequencies or proportions differ significantly among themselves. The data can be nominal or dichotomized ordinal. Cochran's Q-test is an N-samples extension of the McNemar test for the significance of changes in two related samples. This test is useful when a group of individuals has been tested at least three times, and binary data have been collected to characterize a trait or attribute (Sokal and Rohlf 1995). The binary data are coded as 1s and 0s and are analyzed as a modified two-way analysis of variance for a stratified design. A situation where a Q-test would be appropriate would be the measurement of leaf chlorophyll concentrations (low, high) for 30 purple Balduina (*Balduina atropurpurea*) plants in 3 time periods (early, mid-, and late summer).

Mann-Whitney U test (MWUT). MWUT is the most powerful and useful nonparametric alternative to the t-test (Siegel 1956). It is useful when parametric assumptions are strongly violated or the data are ordinal. MWUT possesses greater statistical power than the t-test when the parametric assumption of normality is violated (Connover 1980).

Kolmogorov-Smirnov (K-S) two-sample test. The K-S two-sample test is the nonparametric equivalent of a t-test for comparing two means. The K-S essentially tests two cumulative distributions to assess if they are statistically homogeneous, and is sensitive to distribution differences in location (i.e., means), dispersion, and skewness (Siegel 1956). The test is used to determine if two populations are equivalent with respect to some measured characteristic.

Kolmogorov-Smirnov goodness-of-fit test. This K-S test compares the distribution of a sample with a known frequency distribution and determines if the two distributions are significantly different.

Kruskal-Wallis test. The Kruskal-Wallis test is a one-way analysis of variance by ranks. It is the nonparametric equivalent of one-way ANOVA, and is used to test if N independent samples are from different populations or the populations are homogeneous (the null hypothesis).

Spearman rho and Kendall tau b rank correlation tests. These correlation tests evaluate the degree of association or correlation between two independent variables measured on an ordinal scale. They represent the nonparametric equivalent of the parametric Pearson's correlation coefficient.

Wilcoxon signed-rank test. The signed-rank test is a nonparametric test for making paired comparisons between two variables.

Computer Intensive Methods (CIM)

Computer intensive procedures include a heterogeneous class of statistical techniques, some of which are closely related while others are completely unrelated. Their unifying theme is that they require extensive computer power, and have only recently become popular with the development of economical high speed microcomputers. CIM include: Monte Carlo resampling methods, the calculation of exact P-values (parametric and nonparametric), jackknifing and bootstrapping, permutation tests, and randomization tests. Important references include Miller 1974, Efron 1982, Edgington 1987, Noreen 1989, Efron and Tibshirani 1991, Manly 1991, Shao and Tu 1995, Weerahandi 1995. CIM can also be used for multiple comparisons (Westfall and Young 1993). These techniques are particularly useful for "messy data" (e.g., Milliken and Johnson 1984, 1989), which include: small sample sizes, unbalanced data (dramatic differences in sample sizes of comparisons), strongly skewed data, heterogeneity in residuals, data possessing strange distributions, missing observations, and outliers. For a practical application in the use of CIM for population monitoring of a threatened species (desert tortoise), see Krzysik (1997, 1998a).

Researchers are generally unaware that both parametric and nonparametric tests in a fundamental way rely on asymptotic behavior, which requires reasonable sample sizes and balanced data (Krzysik 1998a). Asymptotic theory is not valid for data sets that are small, highly skewed, sparse, or unbalanced. Statisticians have been aware of the dilemma. "The difficulty of exact calculations coupled with the availability of normal approximations leads to the almost automatic computation of asymptotic distributions and moments for discrete random variables.....How does one justify them?.....Rigorous answers to [this] question ... require some of the deepest results in mathematical probability theory" (Bishop, Fienberg, and Holland 1975). These limitations have been recognized for quite some time, and Fisher (1935) suggested the use of permutational P-values for randomized experiments. The routine use of permutation methods by researchers directly depends on the availability of economic high-powered microcomputers. Today, it is easy to compute exact permuted P-

values for both nonparametric and parametric tests and thus avoid asymptotic assumptions (Mehta, Patel, and Wei 1988; Agresti, Mehta, and Patel 1990; Good 1994). For a discussion of jackknifing and bootstrapping see Krzysik (1998a).

Multivariate Methods

The definition of multivariate statistical methods has not been consistent in textbooks or in the technical literature. In the general sense, multivariate statistics refer to a large body of techniques that deal with the analysis, relationships, and interpretation of multiple-variable data sets. In its most liberal definition, multivariate analysis is the analysis of **more** than two variables. This contrasts with univariate analyses which, in their simplest form, consist of either one dependent and one independent variable (simple linear regression) or two independent variables (bivariate correlation).

Multiple regression involves two or more independent variables, but only one dependent variable. Is this a univariate or multivariate technique? Differences in usage can vary from one source to another. If multivariate analysis is the measure, interpretation, and prediction of the relationships among multiple weighed combinations of variables (variates), then multiple regression is a multivariate technique. It is less ambiguous, however, and used as such for this report, to reserve multivariate terminology for situations involving more than one dependent variable. Multivariate statistics can therefore be defined as the analysis and exploration of data sets containing two or more independent variables and two or more dependent variables. Comparable to the univariate case, parametric assumptions are analogous: multiple variables are assumed to have a multivariate normal distribution, variance and covariance matrices are assumed to be homogeneous, and the multiple variables possess independent errors. Excellent introductions to multivariate analyses include: Pielou (1984), Manly (1986), Digby and Kempton (1987), James and McCulloch (1990), and Marcoulides and Hershberger (1997). For additional references, applications to military training effects, and ecological assessment and monitoring, see Krzysik (1987, 1998a).

Multivariate Analysis of Variance (MANOVA)

MANOVA is used to simultaneously assess the relationships between one or more treatments (independent variables) and two or more dependent variables. Using the minnow example cited above, the dependent variable was survivorship. Another dependent variable that could have been added into the experiment is respiratory rate, making the analysis a three-factor MANOVA.

MANOVA is discussed under repeated-measures analysis in the earlier section on Analysis of Variance (ANOVA) under **Univariate Statistics**.

Multivariate Analysis of Covariance (MANCOVA)

The MANCOVA procedure is similar to ANCOVA, except that more than one dependent variable is under consideration. See the earlier section on Analysis of Covariance (ANCOVA) under **Univariate Statistics**.

Canonical Correlation Analysis (CCA)

Multiple correlation analysis describes the relationships among linear combinations of two or more variables with another single variable. CCA is an analogous technique when there are two sets of two or more variables. As in bivariate correlation, the variables are symmetric, with no assignment of predictor or criterion designations. Assuming that there are n variables in both variable sets x and y , CCA involves finding n linear combinations of x variables (canonical variates or scores, vector X) and n linear combinations of y variables (canonical variates or scores, vector Y), such that vectors X and Y have maximum correlation.

At first glance this appears to be a very valuable tool for environmental studies, because there are many instances where it would be important to correlate the relationships between two variable sets. Unfortunately, CCA is very sensitive to multivariate parametric assumptions, especially to nonlinearity among the original variables and also among canonical variates (linear combinations of individual variables). Linear relationships among environmental variables are extraordinarily rare and atypical in nature and in environmental processes (Krzysik 1987).

An example of the use of CCA is the measuring of many habitat variables in a number of sampling plots (e.g., biomass and cover of forbs and grasses, shrub density, canopy cover, basal area of trees, substrate texture, and soil parameters). Concurrently at the same sampling plots, data are gathered on the species abundances of birds, small mammals, foliage arthropods, and soil litter invertebrates. There are now two major data sets: one with a large number of descriptive habitat variables and another with a large number of population abundance variables. In theory, CCA could provide the optimal linear combinations of habitat variables and of species abundance variables that would best describe the relationship between the two variables.

Principal Component Analysis (PCA)

PCA “produces newly-derived variables from linear combinations of the original variables (often highly correlated), such that most of the original variance in the original data is expressed in as few as possible new uncorrelated variables; and is a powerful procedure for ordination, data reduction, data transformation, and data standardization” (Krzysik 1998a). The most fundamental mathematical description of a multi-parameter environmental gradient is a principal component solution (Krzysik 1987). Environmental gradients are inherent in all ecological and landscape phenomena. Derived principle components can be used in additional statistical analyses.

Discriminant Analysis (DA)

DA is used to describe the nature and extent of differences among groups in multivariate analysis of variance applications and to classify subjects into groups based on multiple measurements. Classification procedures assign subjects into one of several groups based on common characteristics. The subjects are assigned to the groups based on how closely their individual classification scores resemble the classification score for each group as a whole. For example, T&E species may be placed into groups of low, moderate, and high risk of exposure to S/O based on mathematical scores describing physiological or behavioral characteristics (e.g., adaptation to presence of S/O, nesting behavior, food sources, proximity to S/O training).

DA is a popular multivariate technique because it possesses the potential of quantitatively identifying the relative importance of predictor variables in group classifications. Conversely, on the basis of predictor variables, it can classify measured objects or elements into the groups of a previous classification. DA is often used inappropriately because it is unusually sensitive to assumption violations, particularly to the heterogeneity of group covariance structure (Krzysik 1987). This technique should only be used with caution or by experienced statistical practitioners.

Interpretation and Presentation of Results

When statistical results are reported, information concerning the reliability of parameters should be summarized. The minimum information for analysis results should always include: sample size, standard deviation (sometimes standard errors or confidence intervals are more appropriate), type of analysis procedure used to obtain results, and the computer package and version used to

generate the result (Ellison 1993; Taylor 1990). Additionally, a statistical power analysis should be conducted and the results reported (Cohen 1988).

Stressor-response analysis (U.S. EPA 1992) is used to describe the relationship between the amount, frequency, or duration of a stressor and the magnitude of response. In situations where only a limited number of observations can be taken, or where surrogate species must be used, extrapolation of results may be necessary to estimate the effects over a wider range of conditions than are present in the actual study. Types of extrapolations often used in the context of risk assessments (U.S. EPA 1992) and other studies are:

1. Extrapolation between taxa (e.g., measure response of one species, then extend results to other species)
2. Extrapolation between responses (e.g., measure one level of response (LD_{50}), then extend results to other levels (no observed effect level))
3. Extrapolation from laboratory to field (e.g., measure mouse mortality under laboratory conditions, then extend results to field conditions)
4. Extrapolation from field to field (e.g., conduct study in one training area or ecosystem, extend results to other training areas or ecosystems)
5. Analysis of indirect effects (e.g., relating reduced food or habitat resources to reduced T&E species populations)
6. Analysis of higher organizational levels (e.g., relating survival of individual organisms to population size)
7. Analysis of spatial and temporal scale (e.g., evaluating loss of a specific habitat area to the larger scale habitat requirements of a species)
8. Analysis of recovery (e.g., relating short-term effects of catastrophic events to long-term species survival).

Statistical Significance Versus Ecological Significance

Statistical significance is used to denote whether the data collected support or fail to support a null hypothesis in manipulative research. If the data support the null hypothesis, then the response to the treatment under consideration is considered to be essentially the same as the hypothesized response. If the data fail to support the null hypothesis, then the response to the treatment under consideration is considered to be significantly different from the hypothesized response. The researcher must interpret the results of the study in the context of nature and magnitude of the effects, the spatial and temporal patterns of the effects, the likelihood that the effects will occur in a natural context, and the recovery potential of a system from the effect observed (U.S. EPA 1992).

Biological or ecological significance represents biological realism and common sense directly in the context of actual ecological systems and their inherent vari-

ability and unpredictability. "Statistical significance is only relevant to sample size in the specific context of the probability of finding an observed difference by chance alone relative to the inherent variability in the system under investigation. Biological relevance does not enter into the equation. Statistical significance will *always* be assured as long as sample size is made large enough, to 'statistically detect' even the smallest differences. Differences that are undoubtedly irrelevant to the normal course of biological variability" (Krzysik 1998a).

Lovett (1994) found that short-term studies on atmospheric deposition of pollutants can be misleading because individual portions of the longer-term record considered separately would indicate increases, decreases, or no change. The real trends in the data can be obscured by short-term fluctuations as a result of the extreme variability often found in these kinds of studies.

Relationship Between Statistics and Ecological Risk Assessment

Effects of S/O on T&E species may be most effectively assessed in the context of an ecological risk assessment (Sample et al. 1997). Statistics may be applied in two ways in performing ecological risk assessments: in models for assessment and to quantify uncertainty.

Suter and Barnthouse (1993) discuss methods of assessment applicable to ecological risk assessment, including physical methods (test systems as discussed in Sample et al. 1997) and quantitative methods, both statistical and mathematical. Because there is no universal method for quantifying ecological risks, all having limitations, these methods are often complementary ways to quantify exposures, effects, and risks.

Statistical models attempt to derive generalizations by using statistical techniques, such as ANOVA, regression, or principal components analysis (described earlier) to summarize experimental or observational data (Suter and Barnthouse 1993). Toxicologists, for example, obtain dose-response models by statistically fitting a continuous function such as the probit to the discontinuous results of toxicity tests of discrete doses. Such a model assumes that the sensitivities of exposed organisms to a toxic chemical can be characterized by statistical distribution with a mean and a variance.

Suter and Barnthouse (1993) list three purposes for using statistical models in risk assessment: hypothesis testing, description, and extrapolation. Hypothesis testing has been used in risk assessment to calculate "no effects" concentrations in toxicity tests and comparison of contaminated and reference sites in monitoring studies. Caution is required, however, when using hypothesis testing in risk

assessment because, as stated earlier, statistical significance is not the same as ecological significance. Contaminant concentrations in soil, for example, may average 10 times the average background concentration and may be above phytotoxic levels but still not be "significantly elevated" in strictly statistical terms.

The second use of statistical models is description. For example, a multivariate classification method such as principal component analysis might be used to distinguish the sets of natural and contaminant-affected communities of organisms within an ecosystem.

The third use of statistical models is extrapolation. For example, a concentration-response model of a red-winged blackbird toxicity test that describes the response under laboratory conditions may be extrapolated to red-winged blackbirds in the field, to an endangered bird species with relevant similarities (e.g., red-cockaded woodpecker), or to birds in general. Such extrapolations must usually be applied in the case of endangered species. Data extrapolations require that the assessor either assume the systems being compared respond identically or use some extrapolation model (Suter and Barnthouse 1993).

Strictly speaking, a statistical model does not identify causal relationships between independent and dependent variables but simply summarizes the relationship between the variables. However, assignment of biological or physical meaning to the fitted coefficients allows more interpretive weight (Suter and Barnthouse 1993).

The most important feature distinguishing risk assessment from impact assessment is emphasis on characterizing and quantifying uncertainty (Suter and Barnthouse, 1993). Of particular interest in ecological risk assessment are three types of uncertainty that contribute to "analytical uncertainty," or uncertainty in estimating the credibility of a predicted value (Suter, Barnthouse, and O'Neill 1987). These types are natural stochasticity, parameter error, and model error. The first two types can be quantified using statistical models. Although straightforward in concept, use of statistics to quantify uncertainty is complicated in practice by the need to consider measurement errors in both the dependent and independent variables and to combine errors when multiple extrapolations must be made (Linder 1987).

4 Summary

This report provided a general overview of sampling designs and statistical procedures for assessing the effects of military S/O on T&E species. Important fundamental principles were summarized and documented with extensive literature references to provide more detailed information.

Sampling design considerations and strategies were discussed. Types of sample designs and appropriate conditions for the use of each were identified.

Also discussed were statistical analysis considerations. Data types and characteristics of data quality were identified. Approaches to statistical analysis were identified and discussed. The six general approaches to statistical analysis discussed were estimation, description, exploratory analysis, inference, modeling, and spatial analysis. Specific statistical analysis methods were identified and conditions for the use, as well as cautions and pitfalls to avoid, were described for each method. Univariate and multivariate methods were addressed. Assumptions upon which parametric methods are based were stated, and specific parametric and nonparametric methods discussed.

Guidance was provided for interpretation and presentation of statistical results. Finally, statistical significance versus ecological significance and the relationship between statistics and ecological risk assessment were discussed.

References

3D/International Inc., Environmental Group. 1996. *Environmental Fate of Fog Oil at Fort McClellan, Alabama*. Report submitted to Kansas City District, U.S. Army Corps of Engineers.

Agresti, A., C.R. Mehta, and N.R. Patel. 1990. "Exact inference for contingency tables with ordered categories," *Journal of the American Statistical Association* 85:453-458.

Alados, E.L., J.M. Emlen, B. Wachocki, and D.C. Freeman. 1998. "Instability of development and fractal architecture in dryland plants as an index of grazing pressure," *Journal of Arid Environments* 38:63-76.

Barcelona, M.J. 1988. "Overview of the Sampling Process," in Keith, L.H. (Ed), *Principles of Environmental Sampling*, American Chemical Society, Washington, DC, pp 3-23.

Barcikowski, R.S. and R.R. Robey. 1984. "Decisions in a single group repeated-measures analysis: Statistical tests and three computer packages," *American Statistician* 38:148-150.

Bausum, H.T. and G.W. Taylor. 1986. *A Literature Survey and Data Base Assessment: Microbial Fate of Diesel Fuel and Fog Oils*. TR-8408/ADA167799. U.S. Army Materiel Bioengineering Research and Development Laboratory, Fort Detrick, MD.

Bender, F.E., L.W. Douglass, and A. Kramer (Eds). 1989. *Statistical Methods for Food and Agriculture*, Food Products Press, Inc./Haworth Press, Inc., Binghamton, NY.

Berkson, J. 1942. "Tests of significance considered as evidence," *Journal of the American Statistical Association* 37:325-335.

Beyer, W.H., ed. 1988. *Handbook of Tables for Probability and Statistics*, 2nd ed. CRC Press, Inc., Boca Raton, FL.

Bishop, Y.M.M., S.E. Fienberg, and P.W. Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge.

Borgman, L.E. and W.F. Quimby. 1988. "Sampling for Tests of Hypothesis When Data Are Correlated in Space and Time," in Keith, L.H. (Ed), *Principles of Environmental Sampling*. American Chemical Society, Washington, DC, pp 25-43.

Bowers, J.F. and J.M. White. 1992. "Air Quality Modeling for Smoke/Obscurant Pretest Environmental Assessment," in Gerard, S. and W. Klimek (Eds), *Proceedings of the Smoke/Obscurants Symposium XVI, Volume II*, pp 795-810. Technical Report CRDEC-CR-184. Chemical Research, Development, and Engineering Center (CRDEC), Aberdeen Proving Ground, MD.

Box, G.E.P. 1949. "A general distribution theory for a class of likelihood criteria," *Biometrika* 36:317-346.

Box, G.E.P., W.G. Hunter, and J.S. Hunter. 1978. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. John Wiley and Sons, New York.

Brown, A.W.A. 1978. *Ecology of Pesticides*. John Wiley and Sons, New York.

Brubaker, K.L., D.H. Rosenblatt, and C.T. Snyder. 1992. *Environmental Effects of Fog Oil and CS Usage at the Combat Maneuver Training Center, Hohenfels, Germany*. ANL/ESD/TM-38. Argonne National Laboratory, Argonne, IL.

Burgman, M.A., S. Ferson, and H.R. Akcakaya. 1993. *Risk Assessment in Conservation Biology*. Chapman and Hall, New York.

Campbell, R.C. 1989. *Statistics for Biologists*, 3rd ed. Cambridge University Press, New York.

Chambers, J.M., W.S. Cleveland, B. Kleiner, and P.A. Tukey. 1983. *Graphical Methods for Data Analysis*. Duxbury Press, Boston.

Chatfield, C. 1988. *Problem Solving: A Statistician's Guide*. Chapman and Hall, New York.

Chatterjee, S. and B. Price. 1991. *Regression Analysis by Example*, 2nd ed. John Wiley and Sons, New York.

Cleveland, W.S. 1993. *Visualizing Data*. AT&T Bell Laboratories. Hobart Press, Summit, NJ.

Cochran, W.G. 1951. "Testing a linear relation among variances," *Biometrics* 7:17-32.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

Cohen, J., and P. Cohen. 1983. *Applied Regression / Correlation Analysis for the Behavioral Sciences*, 2nd ed. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

Connell, J.H. and W.P. Sousa. 1983. "On the evidence needed to judge ecological stability or persistence," *American Naturalist* 121:789-824.

Conover, W.J. 1980. *Practical Nonparametric Statistics*, 2nd ed. John Wiley and Sons, New York.

Cox, D.R. 1958. *Planning of Experiments*. John Wiley and Sons, New York.

Crowder, M.J. and D.J. Hand. 1990. *Analysis of Repeated Measures*. Chapman and Hall, New York.

Cullen, M.R. 1985. *Linear Models in Biology*. Ellis Horwood, Chichester, England.

Digby, P.G.N. and R.A. Kempton. 1987. *Multivariate Analysis of Ecological Communities*. Chapman and Hall, New York.

Di Giulio, R.T., W.H. Benson, B.M. Sanders, and P.A. VanVeld. 1995. "Biochemical mechanisms: metabolism, adaptation, and toxicity," in Rand, G.M. (Ed), *Fundamentals of Aquatic Toxicology: Effects, Environmental Fate, and Risk Assessment*, 2nd edition, pp 523-561. Taylor & Francis, Washington, DC.

Dobson, A.J. 1990. *An Introduction to Generalized Linear Models*. Chapman and Hall, New York.

Draper, N.R., and H. Smith. 1981. *Applied Regression Analysis*, 2nd ed. John Wiley and Sons, New York.

Eberhardt, L.L. 1976. "Quantitative ecology and impact assessment," *Journal of Environmental Management* 42:1-31.

Eberhardt, L.L. and J.M. Thomas. 1991. "Designing Environmental Field Studies," *Ecological Monographs* 61:53-73.

Edginton, E.S. 1987. *Randomization Tests*. Marcel Dekker, New York.

Efron, B. 1982. *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia.

Efron, B. and R. Tibshirani. 1991. "Statistical analysis in the computer age," *Science* 253:390-395.

Ellison, A.M. 1993. "Exploratory Data Analysis and Graphic Display," in Scheiner, S.M. and J. Gurevitch (Eds), *Design and Analysis of Ecological Experiments*, pp 14-45. Chapman and Hall, New York.

Farmer, W.M. and R.E. Davis. 1986. "Evaluation of Phosphorus Mass Concentration Data Acquired in Field Tests," in *Proceedings of the Smoke/Obscurant Symposium X, Volume I: Unclassified Section*, pp 55-81. Technical Report AMCPM-SMK-T-001-86. Office of the Project Manager Smoke/Obscurants, Aberdeen Proving Ground, MD.

Fisher, R.A. 1935. *The Design of Experiments*. Oliver & Boyd, Edinburgh, London, England.

Foreman, E.K. 1991. *Survey Sampling Principles*. Marcel Dekker, New York.

Freeman, D.C., J.H. Graham, and J.M. Emlen. 1994. "Developmental stability in plants: symmetries, stress, and epigenetic effects," in Markow, T. (Ed), *Developmental Instability: Origins and Evolutionary Significance*, pp 99-121. Kluwer, Dordrecht, The Netherlands.

Getz, L.L., K.A. Reinbold, D.J. Tazik, T.J. Hayden, and D.M. Cassels. 1996. *Preliminary assessment of the potential impact of fog oil smoke on selected threatened and endangered species*. Construction Engineering Research Laboratory (CERL) Technical Report (TR) 96/38/ADA306219. CERL, Champaign, IL.

Goldberg, D.E. and S.M. Scheiner. 1993. "ANOVA and ANCOVA: Field Competition Experiments," in Scheiner, S.M. and J. Gurevitch (Eds), *Design and Analysis of Ecological Experiments*, pp 69-93. Chapman and Hall, New York.

Good, P. 1994. *Permutation Tests - A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer-Verlag, New York.

Graham, J.H., D.C. Freeman, and J.M. Emlen. 1993. "Developmental instability: A sensitive indicator of populations under stress," in Landis, G., J. Hughes, and M.A. Lewis (Eds), *Environmental Toxicology and Risk Assessment*, ASTM STP 1179, pp 136-158. American Society for Testing and Materials, Philadelphia, PA.

Green, R.A., M.K. Cox, T.B. Doerr, T.F. O'Farrell, W.K. Ostler, K.R. Rautenstrauch, and C.A. Wills. 1991. "Assessing impacts on biological resources from site characterization activities of the Yucca Mountain Project," in *Proceedings of High Level Radioactive Waste Management Conference*. April 28 - May 3, 1991. American Nuclear Society, American Society of Civil Engineers, Reston, VA, pp 1456-1460.

Green, R.H. 1979. *Sampling Design and Statistical Methods for Environmental Biologists*. John Wiley and Sons, New York.

Guelta, M.A. and R.T. Checkai. 1995. *Predictive Ecological Risk Assessment of Graphite Infrared Wavelength Obscurant in a Terrestrial Environment*. ERDEC-TR-240. U.S. Army ERDEC, Research and Technology Directorate, Aberdeen Proving Ground, MD.

Gurevitch, J. and S.T. Chester, Jr. 1986. "Analysis of Repeated Measures Experiments," *Ecology* 67:251-255.

Haines, D. 1993a. *Obscurant Precipitation Survey, Smoke Week XIV*. Technical Report PSL-93/16. Physical Science Laboratory (PSL), New Mexico State University, Las Cruces, NM.

Haines, D. 1993b. *Obscurant Precipitation Survey, Smoke Week XV*. Technical Report PSL-93/53. PSL, New Mexico State University, Las Cruces, NM.

Hairston, N.G., R.W. Hill, and U. Ritte. 1981. "The Interpretation of Aggregation Patterns," in Patil, G.P., E.C. Pielou, and W.E. Waters (Eds), *Spatial Patterns and Statistical Distributions*, pp 337-353. The Pennsylvania State University Press, University Park, PA.

Harris, D.J. 1984. "2,3,7,8-Tetrachlorodibenzo-p-dioxin Sampling Methods," in *Environmental Sampling for Hazardous Wastes*, pp 27-35. American Chemical Society, Washington DC.

Harris, R.J. 1975. *A Primer of Multivariate Statistics*. Academic Press, New York.

Hastings, N.A.J. and J.B. Peacock. 1975. *Statistical Distributions: A Handbook for Students and Practitioners*. John Wiley and Sons, Inc., New York.

Hollander, M., and D.A. Wolfe. 1973. *Nonparametric Statistical Methods*. John Wiley and Sons, New York.

Hurlbert, S.H. 1984. "Pseudoreplication and the Design of Ecological Field Experiments," *Ecological Monographs* 54:187-211.

Iman, R.L. and W.J. Conover. 1983. *A Modern Approach to Statistics*. John Wiley and Sons, New York.

James, F.C. and C.E. McCulloch. 1990. "Multivariate analysis in ecology and systematics: panacea or Pandora's box?" *Annual Review of Ecology and Systematics* 21:129-166.

Jernelov, A., K. Beijer, and L. Soderlund. 1978. "General aspects of toxicology," in Butler, G.C. (Ed), *Principles of Ecotoxicology*, pp 151-168. Scientific Committee on Problems of the Environment (SCOPE) 12. John Wiley and Sons, New York.

Johnson, D.H. 1995. "Statistical sirens: the allure of nonparametrics," *Ecology* 76:1998-2000.

Kachigan, S.K. 1986. *Statistical Analysis: An Interdisciplinary Introduction to Univariate & Multivariate Methods*. Radius Press, New York.

Keith, L.H. 1991. *Environmental Sampling and Analysis: A Practical Guide*. Lewis Publishers, Chelsea, MI.

Kendall, R.J. and T.E. Lacher, Jr. (Eds). 1994. *Wildlife Toxicology and Population Modeling: Integrated Studies of Agroecosystems*. Lewis Publishers, Boca Raton, FL.

Kent, M. and P. Coker. 1992. *Vegetation Description and Analysis: A Practical Approach*. John Wiley and Sons, New York.

Korte, F., W. Klein, and P. Sheehan. 1985. "The Role and Nature of Environmental Testing Methods," in Sheehan, P., F. Korte, W. Klein, and P. Bordeau (Eds), *Appraisal of Tests to Predict the Environmental Behavior of Chemicals*. Scientific Committee on Problems of the Environment (SCOPE) 25. John Wiley and Sons, New York.

Krebs, C.J. 1989. *Ecological Methodology*. Harper and Row, New York.

Krzysik, A.J. 1984. "Habitat relationships and the effects of environmental impacts on the bird and small mammal communities of the central Mojave Desert," in McComb, W.C. (Ed), *Proceedings - Workshop On Management of Nongame Species and Ecological Communities*, pp 358-394. University of Kentucky, Lexington, KY.

Krzysik, A.J. 1985. *Ecological assessment of the effects of Army training activities on a desert ecosystem: National Training Center, Fort Irwin, California*. CERL TR N-85/13/ADA159248. CERL, Champaign, IL.

Krzysik, A.J. 1987. *Environmental gradient analysis, ordination, and classification in environmental impact assessments*. CERL TR N-87/19/ADA187294. CERL, Champaign, IL.

Krzysik, A.J. 1994. *Biodiversity and the threatened/endangered/sensitive species of Fort Irwin, California: The National Training Center mission, training effects, and options for natural resources management and mitigation*. CERL TR EN-94/07/ADA291289. CERL, Champaign, IL.

Krzysik, A.J. 1997. "Desert tortoise populations in the Mojave Desert and a half-century of military training activities," in Van Abbema, J. (Ed), *Proceedings: Conservation, Restoration, and*

Management of Tortoises and Turtles -- An International Conference, pp 61-73; July 1993, State University of New York, Purchase. New York Turtle and Tortoise Society, New York.

Krzysik, A.J. 1998a. "Ecological design and analysis: Principles and issues in environmental monitoring," in Lannoo, M.J. (Ed), *Status and Conservation of Midwestern Amphibians*, pp 385-403. University of Iowa Press, Iowa City, IA.

Krzysik, A.J. 1998b. "Geographic information systems, landscape ecology, and spatial modeling," in Lannoo, M.J. (Ed), *Status and Conservation of Midwestern Amphibians*, pp 404-428. University of Iowa Press, Iowa City, IA.

Krzysik, A.J. 1999. *Assessment and Monitoring of Threatened, Endangered, and Rare Populations: Foundations*. CERL TR (draft). CERL, Champaign, IL.

Kuhn, T. 1970. *The Structure of Scientific Revolutions*, 2nd ed. University of Chicago Press, Chicago, IL.

Landis, W.G., G.B. Matthews, R.A. Matthews, and A. Sergeant. 1994. "Application of Multivariate Techniques to Endpoint Determination, Selection, and Evaluation in Ecological Risk Assessment," *Environmental Toxicology and Chemistry* 13(12):1917-1927.

Landrum, P.F., G.A. Harkey, and J. Kukkonen. 1996. "Evaluation of organic contaminant exposure in aquatic organisms: the significance of bioconcentration and bioaccumulation," in Nowman, M.C. and C.H. Jagoe (Eds), *Ecotoxicology: A Hierarchical Treatment*, pp 85-131. CRC Press, Inc.

Lessler, J.T. and W.D. Kalsbeek. 1992. *Nonsampling Error in Surveys*. John Wiley and Sons, New York.

Levene, H. 1960. "Robust tests for equality of variance," in Olkin, I. (Ed), *Contributions to Probability and Statistics*, pp 278-292. Stanford University Press, Palo Alto, CA.

Liljegren, J.C., W.E. Dunn, G.E. DeVaul, and A.J. Policastro. 1988. *Field Measurement and Model Evaluation Program for Assessment of the Environmental Effects of Military Smokes: Field Study of Fog-Oil Smokes*. U.S. Army Medical Research and Development Command, Fort Detrick, MD.

Liljegren, J.C., W.E. Dunn, G.E. DeVaul, and A.J. Policastro. 1989. *The Atterbury-87 Field Study of Smoke Dispersion and a New Stochastic Dispersion Model*. U.S. Army Medical Research and Development Command, Fort Detrick, MD.

Linder, E. 1987. *Statistical Inference in the Linear Errors-in-Variables Model Using the Bootstrap, With Applications in Environmental Risk Analysis*. The Pennsylvania State University, University Park, PA.

Lovett, G.M. 1994. "Atmospheric Deposition of Nutrients and Pollutants in North America: An Ecological Perspective," *Ecological Applications* 4(4):629-650.

MacKay, D., L.A. Burns, and G.M. Rand. 1995. "Fate Modeling," in Rand, G.M. (Ed), *Fundamentals of Aquatic Toxicology: Effects, Environmental Fate, and Risk Assessment*, 2nd ed, pp 563-586. Taylor & Francis, Washington, DC.

Manly, B.F.J. 1986. *Multivariate Statistical Methods: A Primer*. Chapman and Hall, New York.

Manly, B.F.J. 1991. *Randomization and Monte Carlo Methods in Biology*. Chapman and Hall, New York.

Marcoulides, G.A. and S.L. Hershberger. 1997. *Multivariate Statistical Methods: A First Course*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

Maxwell, S.E. and H.D. Delaney. 1990. *Designing experiments and analyzing data: A model comparison perspective*. Wadsworth, Belmont, CA.

McCullagh, P. and J.A. Nelder. 1983. *Generalized Linear Models*. Chapman and Hall, New York.

Mead, R. 1988. *The Design of Experiments: Statistical Principles for Practical Application*. Cambridge University Press, New York.

Mehta, C.R., N.R. Patel, and L.J. Wei. 1988. "Computing exact significance tests with restricted randomization rules," *Biometrika* 75:295-302.

Miller, R.G. 1974. "The jackknife - a review," *Biometrika* 61:1-15.

Milliken, G.A. and D.E. Johnson. 1984. *Analysis of Messy Data, Volume I: Designed Experiments*. Wadsworth, Belmont, CA.

Milliken, G.A. and D.E. Johnson. 1989. *Analysis of Messy Data, Volume II: Nonreplicated Experiments*. Van Nostrand Reinhold, New York.

Møller, A.P. and J.P. Swaddle. 1997. *Asymmetry, Developmental Instability and Evolution*. Oxford University Press, New York.

Montgomery, D.C. and E.A. Peck. 1982. *Introduction to Linear Regression Analysis*. John Wiley and Sons, New York.

Motulsky, H. 1995. *Intuitive Biostatistics*. Oxford University Press, New York.

Myers, J.L. and A.D. Well. 1995. *Research Design and Statistical Analysis*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

Myers, R.H. 1986. *Classical and Modern Regression with Applications*. Prindle, Weber, and Schmidt Publishers, Boston, MA.

Nam, Sae-Im, Marianne E. Walsh, Jean M. Day, and Keturah A. Reinbold. 1999. *Methods for Field Studies of the Effects of Military Smokes, Obscurants, and Riot-control Agents on Threatened and Endangered Species, Volume 4: Chemical Analytical Methods*, TR 99/56/ADA368051. CERL, Champaign, IL.

National Research Council (NRC). 1981. *Testing for Effects of Chemicals on Ecosystems*. National Academy Press, Washington, DC.

Neter, J., W. Wasserman, and M.H. Kutner. 1985. *Applied Linear Statistical Models: Regression, Analysis of Variance and Experimental Design*, 2nd ed. Richard D. Irwin, Homewood, IL.

Noether, G.E. 1987. "Sample size determination for some common nonparametric tests," *Journal of the American Statistical Association* 82:645-647.

Noreen, E.W. 1989. *Computer Intensive Methods for Testing Hypotheses: An Introduction*. John Wiley and Sons, New York.

Noss, R.F., and A.Y. Cooperrider. 1994. *Saving Nature's Legacy: Protecting and Restoring Biodiversity*. Island Press, Washington, DC.

Palmer, A.R. and C. Strobeck. 1986. "Fluctuating asymmetry: Measurement, analysis, patterns," *Annual Review of Ecology and Systematics* 17:391-421.

Passivirta, J. 1991. *Chemical Ecotoxicology*. Lewis Publishers, Inc., Chelsea, MI.

Pielou, E.C. 1984. *The Interpretation of Ecological Data: A Primer on Classification and Ordination*. John Wiley and Sons, New York.

Policastro, A.J., D.M. Maloney, W.E. Dunn, and D.E. Brown. 1991. *Evaluation of Atmospheric Dispersion Models for Smoke Dispersion*. U.S. Army Medical Research and Development Command, Fort Detrick, MD.

Potvin, C. 1993. "ANOVA: Experiments in Controlled Environments," in Scheiner, S.M. and J. Gurevitch (Eds), *Design and Analysis of Ecological Experiments*, pp 46-68. Chapman and Hall, New York.

Potvin, C. and D.A. Roff. 1993. "Distribution-free and robust statistical methods: viable alternatives to parametric statistics," *Ecology* 74:1617-1628.

Racine, C.H., M.E. Walsh, B.D. Roebuck, C.M. Collins, D. Calkins, I. Reitsma, P. Buchli, and G. Goldfarb. 1992. "White phosphorus poisoning of waterfowl in an Alaskan salt marsh," *Journal of Wildlife Disease* 28:669-673.

Riha, J. 1988. "Total index of environmental quality as applied to water resources," in Richardson, M.L. (Ed), *Risk Assessment of Chemicals in the Environment*, pp 363-378. The Royal Society of Chemistry, London.

Sample, B.E., T.L. Ashwood, B.A. Carrico, L.A. Kszos, M.S. Nazerias, T.L. Phipps, W.K. Roy, M.G. Ryon, E.M. Schilling, M. Smith, A.J. Stewart, G.W. Suter II, L.F. Wicker, and K.A. Reinbold. 1997. *Methods for Field Studies of Effects of Military Smokes, Obscurants, and Riot-Control Agents on Threatened and Endangered Species, Volume 2: Methods for Assessing Ecological Risks*. CERL TR 97/140/ADA333828. CERL, Champaign, IL.

Sandusky, M.C. 1992. *Guide for Collecting, Handling, and Preserving Soil Samples for the Analysis of Chemical Agents*. Technical Report CRDEC-CR-143. CRDEC, Aberdeen Proving Ground, MD.

Shao, J., and D. Tu. 1995. *The Jackknife and Bootstrap*. Springer-Verlag, New York.

Shapiro, S.S., M.B. Wilk, and H.J. Chen. 1968. "A comparative study of various tests for normality," *Journal of the American Statistical Association* 63:1343-1372.

Shaw, R.G. and T. Mitchell-Olds. 1993. "ANOVA for unbalanced data: an overview," *Ecology* 74:1638-1645.

Shinn, J.H., L. Sharmer, and M. Novo. 1987. *Smokes and Obscurants: A Guidebook of Environmental Assessment, Vol II, A Sample Environmental Assessment*. ADA203909. U.S. Army Medical and Research Command, Fort Detrick, MD.

Siegel, S. 1956. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York.

Simini, M. 1992. *Guide for Collecting, Handling, and Preserving Vegetation Samples for the Analysis of Chemical Agents*. Technical Report CRDEC-CR-148. CRDEC, Aberdeen Proving Ground, MD.

Smith, F., S. Kulkarni, L.E. Myers, and M.E. Messner. 1988. In Keith, L.E. (Ed), *Principles of Environmental Sampling*, pp 157-168. American Chemical Society, Washington, DC.

Smith, S.M. 1995. "Distribution-free and robust statistical methods: viable alternatives to parametric statistics?" *Ecology* 76:1997-1998.

Snedecor, G.W. and W.G. Cochran. 1989. *Statistical Methods*, 6th ed. Iowa State University Press, Ames, IA.

Sokal, R.R. and F.J. Rohlf. 1995. *Biometry: The Principles and Practice of Statistics in Biological Research*, 3rd ed. W.H. Freeman and Co., New York.

Steel, G.D. and J.H. Torrie. 1980. *Principles and Procedures of Statistics: A Biometrical Approach*, 2nd ed. McGraw-Hill, New York.

Stevens, J. 1992. *Applied Multivariate Statistics for the Social Sciences*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

Stevens, J. 1996. *Applied Multivariate Statistics for the Social Sciences, 3rd ed., with SPSS for Windows Supplement*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

Stewart-Oaten, A. 1995. "Rules and judgments in statistics: three examples," *Ecology* 76:2001-2009.

Suter, G.W. II. 1999. "Developing conceptual models for complex ecological risk assessments," *Hum. Ecol. Risk Assess.* 5(2):375-396.

Suter, G.W. and L.W. Barnthouse. 1993. "Assessment Concepts," in Suter, G. (Editor and principal author), *Ecological Risk Assessment*, pp 33-47. Lewis Publishers, Inc., Chelsea, MI.

Suter, G.W. II, L.W. Barnthouse, and R.V. O'Neill. 1987. "Treatment of risk in environmental impact assessment," *Environ. Manage.* 11:295-303.

Sutton, D.B. and N.P. Harmon. 1973. *Ecology: Selected Concepts*. John Wiley and Sons, Inc., New York.

Taylor, G.E., Jr., D.W. Johnson, and C.P. Anderson. 1994. "Air Pollution and Forest Ecosystems: A Regional to Global Perspective," *Ecological Applications* 4(4):662-689.

Taylor, J.K. 1990. *Statistical Techniques for Data Analysis*. Lewis Publishers, Chelsea, MI.

Tracy, M., D.C. Freeman, J.M. Emlen, J.H. Graham, and R.A. Hough. 1995. "Developmental instability as a biomonitor of environmental stress: An illustration using aquatic plants and macroalgae," in Butterworth, F.M. (Ed), *Biomonitoring and Biomarkers as Indicators of Environmental Change*, pp 313- 338. Plenum Press, New York.

Trexler, J.C. and J. Travis. 1993. "Nontraditional regression analysis," *Ecology* 74:1629-1637.

Triegel, E.K. 1988. "Sampling Variability in Soils and Solid Wastes," in Keith, L.H. (Ed), *Principles of Environmental Sampling*, pp 385-394. American Chemical Society, Washington, DC.

Tsubaki, Y. 1998. "Fluctuating asymmetry of the oriental fruit fly (*Dacus dorsalis*) during the process of its extinction from the Okinawa Islands," *Conservation Biology* 12:926-929.

Tufte, E.R. 1983. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT.

Tufte, E.R. 1990. *Envisioning Information*. Graphics Press, Cheshire, CT.

Tukey, J.W. 1977. *Exploratory Data Analysis*. Addison-Wesley, New York.

Underwood, A.J. 1991. "Beyond BACI: Experimental designs for detecting human environmental impacts on temporal variations in natural populations," *Australian Journal of Marine and Freshwater Research* 42:569-587.

Underwood, A.J. 1992. "Beyond BACI: the detection of environmental impacts on populations in the real, but variable, world," *Experimental Marine Biology and Ecology* 161:145-178.

Underwood, A.J. 1994. "On beyond BACI: Sampling designs that might reliably detect environmental disturbances," *Ecological Applications* 4:3-15.

Underwood, A.J. 1997. *Experiments in Ecology: Their Logical Design and Interpretation Using Analysis of Variance*. Cambridge University Press, New York.

U.S. Code Annotated, Title 16 (Conservation), 1151-3100. 1985. West Publishing Company, St. Paul, MN.

U.S. Environmental Protection Agency (EPA). 1992. *Framework for Ecological Risk Assessment*. EPA/630/R-92/001. Risk Assessment Forum, Washington, DC.

U.S. EPA. 1998. *Guidelines for Ecological Risk Assessment*. EPA/630/R-95/002F. Risk Assessment Forum, Washington, DC.

von Ende, C.N. 1993. "Repeated-measures analysis: Growth and other time-dependent measures," in Scheiner, S.M. and J. Gurevitch (Eds), *Design and Analysis of Ecological Experiments*, pp 113-137. Chapman and Hall, New York.

Weerahandi, S. 1995. *Exact Statistical Methods for Data Analysis*. Springer-Verlag, New York.

Wentzel, R.S. 1986. *Fate and Effects of Brass Powder on the Environment*. ADB 104703. Chemical Research, Development and Engineering Center, Chemical and Biological Defense Command, Aberdeen Proving Ground, MD.

Westfall, P.H. and S.S. Young. 1993. *Resampling-Based Multiple Testing*. John Wiley and Sons, New York.

Wilkinson, L., G. Blank, and C. Gruber. 1996. *Desktop Data Analysis with SYSTAT*. Prentice Hall, Upper Saddle River, NJ.

Winer, B.J. 1962. *Statistical Principles in Experimental Design*. McGraw-Hill, New York.

Winer, B.J., D.R. Brown, and K.M. Michels. 1991. *Statistical Procedures in Experimental Design*, 3rd ed. McGraw-Hill, New York.

Winner, W.E. 1994. "Mechanistic Analysis of Plant Responses to Air Pollution," *Ecological Applications* 4(4):651-661.

Zar, J.H. 1999. *Biostatistical Analysis*, 4th ed. Prentice Hall, Upper Saddle River, NJ.

Zolman, J.F. 1993. *Biostatistics: Experimental Design and Statistical Inference*. Oxford University Press, New York.

Appendix A: Symbols

α	alpha; Type I error
β	beta; Type II error
CI	confidence interval
CV	coefficient of variation
H_A	alternative hypothesis
H_0	null hypothesis
μ	mu; population mean
μ_0	mu naught; population mean for a given reference population
n	number of observations
ρ	rho; population correlation coefficient
r	sample correlation coefficient
r^2	coefficient of determination
s	sample standard deviation
$s_{\bar{x}}$	sample standard error
s^2	sample variance
σ	omicron; population standard deviation
σ^2	population variance
t	t statistic; value for Studentized t-test
\bar{x}	x bar; sample mean
x_i	x sub I; ith observation in a sample

Appendix B: Glossary of Terms

Accessible population – The experimental or sampling units that are actually measured in order to determine an effect; such units may act as surrogates or substitutes for the true units of interest, when taking measurements on the true units is restricted or prohibited.

Accuracy – The amount of systematic error present in a series of measurements.

Alpha (α) – The significance level for a hypothesis test; the probability of making a Type I error, or the probability of rejecting a true null hypothesis; the probability that a given observation will exceed a given critical value if the data are normally distributed.

Alternative hypothesis (H_A) – A statement that indicates the condition that is expected to be true if the null hypothesis, H_0 , is not true. See *Null hypothesis*.

Ambient chemical concentration – Amount of chemical per unit volume of medium (e.g., air, water, soil) in an open environment.

Analysis of covariance (ANCOVA) – A statistical procedure that encompasses both ANOVA (see below) and linear regression. ANCOVA is used when the means of two or more populations are being compared, but the variable of interest is confounded by another variable that may or may not have the same effect on the populations. This variable is called a covariate and linear regression is used to “adjust” for its influence. One of the important tests of ANCOVA is to assess if the regression lines of the populations under analysis possess similar slopes.

Analysis of variance (ANOVA) – A statistical analysis procedure to compare the means of two or more statistical populations and/or treatments. It can also refer to the examination and exploration of sources of variation in sampled data. In this case, it is usually referred to as Variance Component Analysis.

Beta (β) – The probability of making a Type II error; the probability of failing to reject a false null hypothesis; the probability of predicting that no difference exists when a difference is present.

Bias – Systematic error that is associated with a given measurement process and always has the same sign and magnitude.

Bioaccumulation – The uptake/collection of a chemical from the environment into the body of an organism from all routes of exposure.

Bioconcentration – The uptake of a chemical from water by aquatic organisms.

Biomagnification – The tendency of some chemicals to accumulate to higher concentrations in organisms at higher levels in the food chain through dietary accumulation.

Central tendency – The clustering of a set of measurements about a single value that is located approximately midway through the range of values when they are sorted in numerical order.

Classical statistics – Analytical procedures used with manipulative experimental designs to test hypotheses about the condition of the population; inferential statistics.

Coefficient of determination (r^2) – A measure of the amount of variation in the dependent variable that can be attributed to the independent variable.

Coefficient of variation (CV) – A relative measure of the amount of spread or variability for a set of measurements; defined as the standard deviation divided by the mean.

Completely randomized design – An experimental or sampling design that uses a simple random selection process to choose the experimental or sampling units; each unit has an equal and known probability of being selected for measurement.

Conceptual model – A visual or mathematical aid to demonstrate important relationships and hypothesized interactions between stressors and organisms (e.g., S/O and T&E species or T&E habitats).

Confidence level – The degree of certainty the researcher may place in the process used to generate the results of a statistical test.

Confounding factors – Influences other than the ones being explicitly studied that affect the response of a system.

Continuous data – Numbers with measurement uncertainty associated with them.

Correlation – The strength of the relationship between two variables.

Data quality – The accuracy, reliability, and representativeness of measurements.

Degrees of freedom – The amount of information necessary to completely characterize a dependent variable, expressed as the difference between the number of observations and the number of parameters used to estimate variation in the model.

Dependent variable – The name for a set of values that are indirect estimates of a characteristic or property of the population of interest. For example, if the weights of several individuals in rainbow trout population were estimated from known measurements of their lengths, “Weight” would be the dependent variable, and “Length” would be the independent variable.

Descriptive statistics – Analysis methods used to summarize properties of a population, rather than test hypotheses.

Deterministic model – Assume that conditions in the equations remain fixed and constant (i.e., no statistical uncertainty is included in the model), and may be used to describe parameters associated with basic environmental/T&E species states and processes, such as age structure, population size, reproduction rates, environmental conditions, and population growth.

Discrete number – A number with an exact value; a number with no uncertainty due to measurement error.

Discriminant analysis – A multivariate statistical procedure that evaluates several dependent variables in order to assign the various experimental units to distinctive groups.

Dispersion (chemical) – The movement, diffusion, and dissipation of a substance (e.g., of a gaseous suspension of particles in the air).

Dispersion (statistical) – The degree of variability, scatter, or spread of data, usually around a central value; the extent to which sample data differ from the population parameter of interest.

Distribution (statistical) – An arrangement or pattern of data values around a central value that can be described by mathematical functions called probability density functions.

Estimation model – A set of mathematical equations representing the system of interest that is used to identify variables that contribute to explaining chemical or biological processes and to provide probability estimates for events that affect the system.

Experimental design – The set of plans and instructions by which data are collected; the field layout for a manipulative study.

Experimental error – The variation between two observations due to differences between treatments in a manipulative study.

Experimental unit – The smallest subdivision of experimental material (or area) that can receive a given treatment.

Exploratory data analysis – The evaluation of data by summarizing, graphing, or describing; analysis procedures that do not use hypothesis testing.

Fixed factor effect – Results obtained by conducting an analysis of variance on data taken from samples that represent specific levels of interest deliberately chosen by the researcher.

Homoscedasticity – The occurrence of equal variances among treatment groups.

Hypothesis – A statement of an assumed condition that can be confirmed or refuted by additional testing or observation.

Independent variable – The name for a set of values that are direct measurements of a characteristic or property of the population of interest. For example, if the measured lengths of several individuals in rainbow trout population were used to estimate the weights of the trout, then “Length” would be the independent variable, and “Weight” would be the dependent variable.

Inferential statistics – Statistical analysis procedures that test the validity of a hypothesis.

Kurtosis – The relative departure of a sample distribution from a normal distribution in terms of the relative peakedness or flatness of the distribution in the neighborhood of the mode.

Linear regression – A multivariate extension of correlation analysis in which the strength of the relationships between several variables is assessed.

Manipulative study – An experiment characterized by the application of different treatments to different experimental units; an experiment in which events are manipulated or influenced by the researcher.

Mensurative study – The observation or measurement of intrinsic ecological phenomena. The researcher makes no attempt to manipulate or influence events (i.e., apply a treatment) during the course of the study; instead, time or space are used as treatment variables, and inherent properties of the populations or systems are the features of interest.

Multivariate analysis – The analysis of data consisting of more than two dependent and independent variables.

Multivariate analysis of variance (MANOVA) – Analysis method that evaluates sources of variation in sample data when more than one dependent variable is present.

Nonparametric analysis procedure – Statistical analysis procedure that makes no assumptions about the distribution of the data; procedure for data that does not follow a normal distribution.

Normal distribution – The normal or Gaussian distribution refers to a pattern of data values that is commonly called a bell-shaped curve.

Null hypothesis (H_0) – A formal statement or conjecture to be tested by a statistical analysis procedure. The null hypothesis is often worded so as to indicate that no change has occurred or no difference exists.

Outlier – An observation that deviates substantially from the majority of the observations in a data set.

Parameter – A fixed numerical quantity that describes a characteristic of an entire population. Examples of parameters would be mean, median, mode, standard deviation, variance, correlation coefficient, and other numbers used to summarize data. Parameters are denoted by lower-case Greek characters (e.g., μ , σ , ρ).

Parametric analysis procedure – Hypothesis testing procedure based on the assumption that the data follow a normal (Gaussian) distribution.

Pilot study – A small study conducted prior to the main research effort in order to collect preliminary information, to finalize field sampling methods, and to detect weakness in the sampling design.

Population (ecological) – A group of organisms that are close enough to each other to interbreed (i.e., contribute to a common gene pool).

Population (statistical) – The set of numbers that describes all possible events in a defined universe.

Power ($1 - \beta$) – The probability of not making a Type II error. This determines the ability of the statistical test used to detect a true difference when the sample size and α are specified.

Precision – A measure of mutual agreement among individual measurements of the same property.

Predictive model – Mathematical relationships used to characterize system behavior beyond the range of the data.

Principal component analysis – Analysis method that reduces the number of variables needed to account for variation in the data by recombining the variables into uncorrelated linear combinations.

P-value – A value calculated by an inferential analysis procedure. The P-value is used to determine whether or not to accept the null hypothesis. After a statistical analysis is concluded, the P-value is compared to the preselected α . If the absolute value of the P-value is greater than α , then the null hypothesis is accepted, and the alternative hypothesis is rejected. If $|p|$ is less than α , then the test has failed to accept the null hypothesis and the alternate hypothesis is accepted, “there is a significant difference in the means.”

Qualitative hypothesis – General statement of an assumed condition that denotes the relative change or difference to be detected for an assumed condition. This statement can be confirmed or refuted by additional testing.

Quantitative hypothesis – Specific statement of the exact amount of change or difference to be detected for an assumed condition. This statement can be confirmed or refuted by additional testing.

Random error – The fluctuation of sample values around the true value of the parameter of interest, resulting in nonsystematic differences between the sample value and the true value.

Random factor effect – In an analysis of variance, random factor effects are those differences between experimental units that are randomly selected and represent all conceivable levels for the entire population.

Randomization – The assignment of treatments to experimental units in a manner that ensures that each experimental unit has an equal probability of receiving any given treatment.

Regression analysis – See *Linear regression*.

Rejection criteria – The value for the significance level, α , at which a null hypothesis will not be accepted.

Reliability – Data quality that can be documented, evaluated, and believed.

Repeated measures design – A design in which each experimental or sampling unit is sampled more than once. If the study is manipulative, all units receive all treatments in random sequence. If the study is mensurative, no treatments are applied, but each unit is measured for more than one trait or sampled more than one time.

Replication – The assignment of a complete set of treatments more than once during an experiment.

Representativeness – The degree of similarity between the conditions present in a research study and the true condition of the population of interest.

Robustness – The ability of a statistical analysis procedure to give correct results when underlying assumptions for the procedure are violated.

Sample – A subset of a statistical population.

Sampling design – The set of plans and instructions by which the data are collected; the field layout for a mensurative study.

Sampling protocol – A set of written step-by-step instructions for collecting and measuring samples.

Sampling unit – The design elements that are actually measured.

Sensitivity – The ability of an experimental design to detect true differences if they exist; the inverse of the standard deviation for the difference between two means.

Significance (statistical) – The criteria used to denote whether the data collected support or fail to support a null hypothesis in manipulative research.

Skewness – The relative departure of a sample distribution from a normal distribution in terms of the asymmetry at either tail of the distribution. In other words, one of the two tails of the distribution is more drawn out.

Standard deviation – A weighted measure of distance between the observations in a sample and the sample mean; the square root of the variance for a series of measurements.

Standard error – The standard deviation divided by the square root of the number of observations. It is used to indicate the relative precision of the standard deviation when the measurements come from several sets of observations rather than from individual observations.

Stochastic model – A set of mathematical equations that describe a system of interest and that introduce random or chance fluctuations into the system.

Stratification – The partitioning or subgrouping of a population by known characteristics in order to reduce the variability present.

Student's t-test – Statistical method used to evaluate two populations to determine if a difference exists between them.

Surrogate species – A species used as a substitute for another (e.g., a non-T&E species used as a substitute to estimate the effects of S/O on T&E species).

Systematic error – Deviations from a true value that have the same sign and magnitude.

Systematic sampling – Selection of experimental and sampling units in a pre-determined, nonrandom manner.

Target population – The population about which inferences are to be tested; the population of interest.

Test statistic – The value obtained as a result of conducting a given hypothesis test. For example, t is the test statistic for a Student's t-test.

Treatment – Manipulation of experimental material.

Type I error – The probability of rejecting a true null hypothesis (i.e., concluding from test results that a difference exists, when actually no difference is present).

Type II error – The probability of failing to reject a false null hypothesis (i.e., concluding from test results that no difference exists, when actually a difference is present).

Uncertainty (statistical) – The variability in data due to natural fluctuations.

Univariate analysis – Statistical procedure to summarize information about the distribution of the data when, in its simplest form, only one dependent variable and one or more independent variables are being evaluated.

Variability – The difference between the true value of a parameter and the values of each measurement used to estimate the true value; natural fluctuations in data.

Variable – Number used to characterize sample data.

Variance – A weighted measure of distance between the observations in a sample and the sample mean.

Appendix C: Acronyms

ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
ATEC	U.S. Army Test and Evaluation Command
CCA	Canonical correlation analysis
CERL	Construction Engineering Research Laboratory
CIM	Computer intensive methods
CRD	Completely randomized design
DA	Discriminant analysis
EDA	Exploratory data analysis
EPA	Environmental Protection Agency
HC	Hexachloroethane
IDA	Initial data analysis
LOWESS	Logistic regression and locally weighed scatterplot smoothing (regression)
MANOVA	Multivariate analysis of variance
NPS	Nonparametric statistics
NRC	National Research Council
OSL	Observed significance level
PCA	Principal component analysis
RMD	Repeated measures design
S/O	Smokes and obscurants
SRD	Stratified randomized design
TCDD	2,3,7,8-tetrachlorodibenzo-p-dioxin
T&E	Threatened and endangered (species)
TROG	Total recoverable oil and grease
WP	White phosphorus

Appendix D: Checklist for Implementation of Field Research for Evaluating Effects of Military Smokes and Obscurants (S/O) on Threatened and Endangered (T&E) Species

Installation: _____ **Date:** _____

Point of Contact: _____ **Telephone:** _____

E-mail Address: _____ **Fax:** _____

A. Desired outcome of investigation (examples):

1. Descriptive record of population abundance, distribution, etc.
2. Record changes in population over time
3. Compare two or more populations with each other
4. Quantify site or habitat conditions
5. Delineate current status of ecosystem or population
6. Determine interrelationships between biota or ecosystem and S/O
7. Characterize S/O concentrations, dispersion, deposition
8. Quantify direct effects of S/O on T&E species
9. Other _____

B. Population(s) of concern

1. Ecological population
 - a. T&E species
 - b. T&E species surrogates
 - c. Non-T&E species

2. Habitat

- a. Physical populations
- b. Chemical populations

3. S/O

4. Statistical populations (T&E species or S/O traits to be measured)

C. Parameters of interest (types of statistical summaries — totals, means, medians, variances, extremes, correlations, etc.)

D. Facts already known about the situation or problem

1. T&E species information

- a. Listings of T&E species on the installation
- b. Maps of actual or potential T&E species habitat
- c. Locations of T&E individuals or populations
- d. Identification of critical habitat needs for T&E species
- e. Life history of T&E species
- f. Past and current T&E species population trends
- g. Installation reports/memoranda/publications on T&E species
- h. Other _____

2. T&E surrogate species information

- a. Listings of T&E surrogate species on the installation
- b. Maps of actual or potential T&E surrogate species habitat
- c. Locations of T&E surrogate species populations
- d. Identification of critical habitat needs for T&E surrogates
- e. Life history of T&E surrogates
- f. Past and current T&E surrogate species population trends

- g. Installation reports/memoranda/publications on T&E surrogates
- h. Similarities/differences between T&E species and T&E surrogates
- i. Correlations/extrapolations between T&E species and T&E surrogate species responses to military S/O
- j. Other _____

3. Non-T&E species information

- a. Lists of non-T&E species on the installation
- b. Maps of actual or potential non-T&E species habitat
- c. Locations of non-T&E species populations
- d. Identification of critical habitat needs for non-T&E species
- e. Life history of non-T&E species
- f. Past and current non-T&E species population trends
- g. Installation reports/memoranda/publications on non-T&E species
- h. Other _____

4. Military S/O information

- a. Type of S/O
- b. Known physical properties of S/O (e.g., boiling/freezing point, viscosity, solubility, etc.)
- c. Known chemical composition (e.g., fog oil is a mixture of hydrocarbons generally containing 12 to 20 carbon atoms per molecule)
- d. Method of deployment (e.g., stationary or moving generators, smoke pots, grenades, etc.)
- e. Maps showing where S/O releases occur
- f. Quantity of S/O released per unit time period
- g. Quantity of S/O released per unit area

- h. Season and timing of S/O release
- i. Frequency of S/O release
- j. Duration of S/O release
- k. Intensity of S/O release
- l. Records of past and current S/O use
 - For stationary generators
 - Number and location of generators
 - History of current and past configurations
 - History of use
 - For mobile S/O exercises
 - Number and types of S/O releases in a typical exercise
 - Delineation of areas directly affected by S/O exercises (i.e., immediate maneuver area)
 - Delineation of areas indirectly affected by S/O drift
- m. S/O dispersion patterns
- n. Other _____

5. T&E species and S/O interactions

- a. Identification and ranking of research priorities based on
 - Military activities most restricted by T&E species
 - T&E species population trends in S/O areas
 - Future anticipated use of S/O areas
 - Other _____
- b. Delineation of areas where T&E species populations and S/O training activities coincide

- c. Known or anticipated physiological or behavioral changes in T&E species or T&E surrogate species caused by exposure to S/O (e.g., bioassay results)

- d. Other _____

6. General site information

- a. Terrain maps
- b. Vegetation maps
- c. Digital elevation maps
- d. Soils maps
- e. Other maps
- f. Description of ecosystem
- g. Description of selected microhabitats
- h. Land-use history
- i. Ecological history
- j. Weather data
 - Temperature
 - Relative humidity
 - Wind speed and direction
 - Precipitation
- k. Other _____

7. Supplementary information

- a. Types and quantities of nonmilitary chemicals released on or near the installation (e.g., agricultural fertilizers, herbicides, or pesticides; industrial chemical releases)
- b. Regulatory constraints on military activities and S/O-T&E species research

- c. Labor and financial resources available to conduct S/O research
- d. Data from pilot studies, literature reviews, expert opinion, regulatory agencies, etc.
- e. Other _____

E. Assumptions needed to initiate the investigation

- 1. Statistical assumptions
 - a. Distribution of the data
 - b. Presence or absence of spatial, temporal, or other patterns in the data
 - c. Estimated effects of military or nonmilitary activities that might affect data interpretation
 - d. Limitations of sampling design and methods
- 2. Ecological assumptions
 - a. Estimate of nature and extent of problem
 - b. Species and specific populations likely to be affected by S/O
 - c. Informed estimates needed to replace knowledge gaps about species/populations and S/O.

F. Basic nature of the problem: research, inventory, monitoring, or conformance

G. Temporal nature of the problem: one-time, short-range, or long-range

H. Spatial nature of the problem: local, regional, or global

CERL Distribution

Chief of Engineers
ATTN: CEHEC-IM-LH (2)

SERDP Program Office 22203-1853 (3)

Directorate of Environmental Programs
ATTN: DAIM-ED-N 20310-0600

U.S. Army Environmental Center
ATTN: SFIM-AEC-CDN 21010-5401

U.S. Army Training and Doctrine Command
ATTN: ATBO-SE 23651-5000
ATTN: ATIC-ATMS-LM 23604-5166

U.S. Army Forces Command 30330-1062
ATTN: AFPI-ENE
ATTN: AFOP-TE

U.S. Army Materiel Command 61299-7190
ATTN: AMXEN-U (2)

Chief, Army National Guard Bureau 22204-1382
ATTN: NGB-ARE-C
ATTN: NGB-ARO-TS

Engineer Research and Development Center (Libraries)
ATTN: ERDC, Vicksburg, MS
ATTN: Cold Regions Research, Hanover, NH
ATTN: Topographic Engineering Center, Alexandria, VA

Defense Tech Info Center 22304
ATTN: DTIC-O

REPORT DOCUMENTATION PAGE

*Form Approved
OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) 09-2001		2. REPORT TYPE Final		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Methods for Field Studies of the Effects of Military Smokes, Obscurants, and Riot-control Agents on Threatened and Endangered Species Volume 3: Statistical Methods				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Debra M. Cassels, Anthony J. Krzysik, and Keturah A. Reinbold				5d. PROJECT NUMBER SERDP	
				5e. TASK NUMBER CS-766; CS-507	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Engineer Research and Development Center (ERDC) Construction Engineering Research Laboratory (CERL) P.O. Box 9005 Champaign, IL 61826-9005				8. PERFORMING ORGANIZATION REPORT NUMBER ERDC/CERL TR-01-59	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) SERDP Program Office ATTN: Compliance/Conservation Program Manager 901 N. Stuart Street, Suite 303 Arlington, VA 22203				10. SPONSOR/MONITOR'S ACRONYM(S) SERDP	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES Copies are available from the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161.					
14. ABSTRACT <p>Smokes, obscurants, and riot-control agents constitute a diverse group of chemical compounds that are released into the environment during military training exercises. Concern has been expressed over the use of these compounds and how they may affect threatened and endangered (T&E) species that reside on military installations.</p> <p>This report discusses strategies for developing a statistically sound approach to assessing effects of military smokes, obscurants, and riot-control agents on plant and animal species, including T&E species, and their ecosystems. It provides a general overview of sampling designs and statistical analysis procedures and lays a basic foundation for understanding: (1) principles of experimental design and statistical analysis procedures, (2) strengths and weaknesses of each design and analytical procedure, and (3) ecological and statistical rationale for selection of particular designs.</p> <p>Volume 1 of this report series will be an overview of methods examined in the series, their application, and applicable regulations. Volume 2 (CERL TR 97/140, September 1997) reviews methods for assessing ecological risks. Volume 4 (CERL TR 99/56, July 1999) discusses chemical analytical methods.</p>					
15. SUBJECT TERMS threatened species smokes and obscurants military training ecosystem management endangered species riot-control agents Strategic Environmental Research and Development Program (SERDP)					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 104	19a. NAME OF RESPONSIBLE PERSON Keturah A. Reinbold
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (include area code) (217) 352-6511 x6711